8-7-2020

# Applications of Random Effects in Dependent Compound Risk Models

Himchan Jeong
*University of Connecticut - Storrs*, himchan.jeong@uconn.edu

# Applications of Random Effects in Dependent Compound Risk Models

Himchan Jeong, Ph.D.
University of Connecticut, 2020

## ABSTRACT

In the ratemaking for general insurance, calculation of the pure premium has traditionally been based on modeling frequency and severity separately. It has also been a standard practice to assume, for simplicity, the independence of loss frequency and loss severity. However, in recent years, there is a sporadic interest in the actuarial literature and practice to explore models that depart from this independence assumption. Besides, because of the short-term nature of many lines of general insurance, the availability of data enables us to explore the benefits of using random effects for predicting insurance claims observed longitudinally, or over a period of time.

This thesis advances work related to the modeling of compound risks via random effects. First, we examine procedures for testing random effects using Bayesian sensitivity analysis via Bregman divergence. It enables insurance companies to judge whether to use random effects for their ratemaking model or not based on observed data. Second, we extend previous work on the credibility premium of compound sum by incorporating possible dependence as a unified formula. In this work, an informative dependence measure between the frequency and severity components is introduced which can capture both the direction and strength of possible dependence. Third, credibility premium with GB2 copulas are explored so that one can have a succint closed form of the credibility premium with GB2 marginals and explicit approximation of credibility premium with non-GB2 marginals. Finally, we extend microlevel collective risk model into multi-year case using the shared random effect. Such framework includes many previous dependence models as special cases and a specific example is provided with elliptical copulas. We develop the theoretical framework associated with each work, calibrate each model with empirical data and evaluate model performance with out-of-sample validation measures and procedures.

# Applications of Random Effects in Dependent Compound Risk Models

Himchan Jeong

B.B.A., Seoul National University, Republic of Korea, 2012

B.Sc., Seoul National University, Republic of Korea, 2012

M.Sc., Seoul National University, Republic of Korea, 2016

A Dissertation
Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy
at the
University of Connecticut
2020

## APPROVAL PAGE

Doctor of Philosophy Dissertation

# Applications of Random Effects in Dependent Compound Risk Models

Presented by

Himchan Jeong, B.B.A., B.Sc., M.Sc.

Major Advisor ________________________________________

Emiliano A. Valdez

Associate Advisor ________________________________________

Guojun Gan

Associate Advisor ________________________________________

Dipak K. Dey

University of Connecticut
2020

ACKNOWLEDGEMENTS

First of all, I want to glorify my God with my Ph.D. degree and my career in academia which begins with this degree.

I am really blessed to have Professor Emiliano A. Valdez as my Major Advisor. He really has cared for me not only as an academic advisor, but also in personal ways. I learned from him how to take care of the advisees and provide them opportunities to grow as independent researchers.

Professor Guojun Gan supported me with his expertise in programming and writing. I really admire his productivity based on his work ethics and dilligence.

Professor Dipak K. Dey allowed me to ponder the fundamental of Bayesian statistics and its possible applications in actuarial science.

I also thank my collaborators in Korea, Professor Jaeyoun Ahn, Professor Sojung Park, and Dr. Rosy Oh for allowing me to use the collaborations with them as a part of this thesis.

Besides, I appreciate the professors in our department including but not limited to Xiaodong Yan, Damin Wu, Jeyaraj Vadiveloo, and Bin Zou who had provided me their knowledge and supports.

While I have been working on this thesis, I reallized that I am standing on the shoulder of giants, who have set foundations for applications of random effects in insurance ratemaking. Among those giants, especially I want to express my honor to Professors Jose Garrido and Edward W. Frees, who have provided me helpful and kind feedback whenever I meet them.

I also would like to thank my colleagues, actuarial Ph.D. students and alumni at UConn - Jiatian Xu, Zhiyu Quan, Zhiguo Wang, Banghee So, Jinkai Xu, and Yingzhou He - who have shared the same concerns and had good fellowship with me in Monteith 114.

I shall acknowledge that there have been many hands of help in the form of financial aid. I do not believe that I could finish my Ph.D. without such supports which include tuition waiver and stipends from UConn, travel support from SOA CAE and CKER grants, SOA James C. Hickman Doctoral Scholarship, and especially personal help from David Rhee which was really a big help when I was struggling with financial distress.

I cannot forget warm-hearted support and prayer from my family in Korea, Sangjin Jeong, Jaesoon Jeong, and Haneul Jeong.

Last not but least, I want to confess my sincere heart to my lovely wife, Seulgi Choi, who has trusted and supported me without any hesitation.

# Contents

# Chapter 1

# Motivation and literature review

Insurance products are unique commodities because the true cost and profitability are usually difficult to predict at time of sale. Ratemaking in general insurance involves this challenging process of determining a fair, equitable, and reasonable premium suitable for a given class of risk and for a given unit of exposure (e.g., per year, per coverage amount). Premium is a function of the cost arising from claims, expenses, and profits. It is natural to assume that expected cost for each policyholder varies according to specific policyholder characteristics and policy features, which motivates insurance companies to use the concept of regression modeling to develop a mechanism for risk classification. Although linear models have been used as a primitive method of regression which assumes normally distributed random error, it may not be suitable to apply ordinary linear regression for insurance claims because claims are inherently greater than or equal to zero while normally distributed random error could be negative. Therefore, generalized linear model (GLM) was proposed, which extends ordinary linear model with more flexibility. Unlike ordinary linear model, GLM allows the response variable $Y$ to follow any distribution in the exponential family and mean of $Y$ is expressed as follows:

$$\mathbb{E}\left[Y|\mathbf{x}\right] = \mu = g^{-1}(\mathbf{x}\beta),$$

where $\mathbf{x}$ is independent variables (covariates), $\beta$ is the regression coefficient associated with $\mathbf{x}$, and $g$ is link function which connects mean of $Y$ and linear predictor $\mathbf{x}\beta$.

For many apparent reasons including ease of implementation, there has been an increase in popularity even among practitioners of the use of generalized linear models (GLMs) for insurance ratemaking, risk classification and many other actuarial applications. See, for example, Antonio and Valdez (2012), Frees et al. (2014a) and Frees et al. (2016a). Originally synthesized by Nelder and Wedderburn (1972), GLMs extend the ordinary regression models to accommodate response variables that are not normally distributed and are rather members of the exponential family of distributions. As pointed out in Chapter 5 of Frees et al. (2014a), the primary features of GLMs include a function that links the response variable to a set of predictor variables and a variance structure that is not necessarily constant across independent observations. It encompasses a wide variety of popular models that include the normal regression, Poisson regression, logistic regression, probit regression, to name a few.

For a typical portfolio of insurance policies, it is not uncommon to have observations of independent policyholders to come in a longitudinal format such as

$$(N_{it}, C_{itj}, \mathbf{x}_{it}, e_{it})' \tag{1.1}$$

for calendar year $t$, for $t = 1, \ldots, T_i$ where $T_i \leq T$ and for policyholder $i$, for $i = 1, \ldots, M$. There is a fixed number of calendar years $T$ and we allow for unbalanced data. $\mathbf{x}_{it}$ refers to the vector of covariates describing policyholder characteristics and $e_{it}$ refers to the length of exposure of the policyholder within calendar year $t$ where $0 < e_{it} \leq 1$. $C_{it}$, the average severity component is defined as follows:

$$C_{it} = \begin{cases} \dfrac{1}{N_{it}} \sum_{k=1}^{N_{it}} Y_{itk}, & N_{it} > 0 \\ \text{Undefined}, & N_{it} = 0 \end{cases}$$

and $S_{it} = \sum_{k=1}^{N_{it}} Y_{itk} = N_{it}C_{it}$ is the observed aggregate claim size.

Note that the joint density of the number of claims and the average claim size can be decomposed as $f(N, C|\mathbf{x}) = f(N|\mathbf{x}) \times f(C|N, \mathbf{x})$. We have some issues in compound risk model. First, it is easy to see that most of general insurance claim datasets show longitudinal property so that the companies can observe the same policyholder repeatedly over a period of time. Therefore, a company may use random effects for capturing the unobservable heterogeneity of the policyholders via random effects in practice.

One can see that random effects models have been widely used in actuarial science in the framework of credibility theory, which was firstly introduced by Bailey (1950). Afterwards, Mayerson (1964) and Jewell (1974) provided a theoretical discussion on credibility theory from a Bayesian perspective. While these previous research on credibility theory provided a way to incorporate the unobservable heterogeneity, both the observed and unobserved heterogeneity need to be addressed in insurance ratemaking practice as discussed in Norberg (1986). In this regard, Frees et al. (1999) provided a general framework which integrates well-known credibility theory and regression analysis based on the use of linear mixed models. A linear mixed model is an extension of linear model, whereby the response variable is affected by both the observed covariates and associated regression coefficients (fixed effects) and unobserved quantities (random effects). Linear mixed models have been widely used to analyze longitudinal data including but not limited to econometrics, risk management, finance, and biomedical sciences. For instance, Torre et al. (2011) applied linear mixed models to analyze a panel data of income inequality and population health on 21 developed countries. Random effects are also used in Jaba et al. (2017) to control temporal effects on the financial performance assessments of companies. Gurrin et al. (2001) used linear mixed models to analyze foetal growth and control the random fluctuation by each pregnant woman. Mestiri and Hamdi (2012) also used mixed model in credit risk prediction modeling to reflect the unobserved heterogeneity of each company.

The idea of linear mixed models, which enable us to capture both the observed and unobserved heterogeneity in risks of the policyholders, has also been used to construct an effective bonus-malus system (BMS), for example, in Pinquet (1997), Pinquet (1998), and Gómez-Déniz et al. (2005). As extensions of these work, Pinquet et al. (2001) and Bolancé et al. (2003) tried to capture possible evolution in the unobserved heterogeneity and proposed the use of dynamic random effects to update BMS scale of each policyholder over time. Frangos and Vrontos (2001) also used conjugate random effects to derive the credibility premium of the compound sum, which can be expressed as a product of the credibility premiums of the frequency and severity components. Antonio and Beirlant (2007) and Antonio and Valdez (2012) utilized generalized linear mixed models (GLMMs), which is the combination generalized linear models (GLMs) and random effects in order for a posteriori ratemaking. Hernández-Bastida et al. (2009) and Oh et al. (2020) proposed another way to derive the credibility premium of the compound sum, by using bivariate random effects for the frequency and severity components, respectively. Baumgartner et al. (2015) used a shared random effects model, where the same unobserved heterogeneity affects the risk profile of a policyholder both in frequency and severity.

However, even though there has been some important previous work on the use of random effects model in actuarial science literature, a theoretical approach has not been attempted for testing the presence of random effects in claim modeling. Furthermore, although independence between frequency and severity components has been assumed for modeling compound loss for most of the research works aforementioned, recent research works provide empirical evidence of dependence between frequency and severity components, including but not limited to Boudreault et al. (2006), Hernández-Bastida et al. (2009), and Garrido et al. (2016). Finally, there has been no previous research work which considered possible dependence among individual severities in a longitudinal setting.

Therefore, this thesis incorporates these three issues in the ratemaking with compound risk model in the following manner. Chapter 2 introduces a theoretical framework to test the presence of random effects via Bayesian sensitivity analysis. Chapter 3 provides a unified approach which considers both random effects and effects of frequency on average severity with closed form of credibility premium formula. Chapter 4 suggests a change of measure technique to derive credibility premium based on the family of GB2 copulas, which is also constructed via the use of random effects. Chapter 5 introduces a microlevel multi-year model which incorporates possible dependence between frequency and severity as well as among individual severities in a longitudinal setting via shared random effects. Finally, Chapter 6 explores possible directions of future research works with the applications of random effects in dependent compound risk models.

Note that we use different measures for in-sample model selection and out-of-sample validation. For in-sample model selection procedure, we use Akaike information criterion (AIC) Bayesian information criterion (BIC), which are defined as follows:

$$\text{AIC} : -2\ell\left(\hat{\theta}|y\right) + 2p, \quad \text{BIC} : -2\ell\left(\hat{\theta}|y\right) + p\log n$$

where $p$ denotes the number of parameters used in model calibration and $n$ denotes the number of observations used in the calibration. Ideally, we want to maximize the value of (log)likelihood but also consider the model complexity in terms of the number of estimated parameters. Thus, models with lower AIC and BIC are preferred.

Root-mean-square error (RMSE) and mean absolute error (MAE) are used as the out-of-sample validation measures throughout this paper, which are defined as follows:

$$\text{RMSE:} \ \sqrt{\frac{1}{J}\sum_{j=1}^{J}(y_j-\hat{y}_j)^2}, \qquad \text{MAE:} \ \frac{1}{J}\sum_{j=1}^{J}|y_j-\hat{y}_j|$$

where $y_j$ is an actual observed value in the validation set and $\hat{y}_j$ is a predicted value based on the estimated parameters and corresponding covariates $\mathbf{x}_j$. Since both RMSE and MAE measure the discrepancy between the set of actual values and predicted values, models with lower RMSE and MAE are preferred.

# Chapter 2

# Testing the presence of random effects via Bayesian sensitivity analysis

[1] Traditionally, generalized linear models (GLMs) have been used as benchmarks in ratemaking of property and casualty (P&C) companies due to their interpretability and efficiency in modeling. In ratemaking with GLM, regression coefficients associated with the observable characteristics of policyholders (in other words, covariates) are estimated and used for future prediction of claims. But it is not possible to observe all the characteristics of policyholders which affect their risk profiles, such as driving habits. Since a policyholder can be observed repeatedly for many years by a P&C insurance company, one can try to capture the unobserved heterogeneity via random effects model.

Suppose we have the following information on policyholders A and B in Table 2.1, who are identical in terms of observable characteristics but show quite different patterns on their claims. This hypothetical example shows us that we might capture the unobserved heterogeneity in risk by observing the residuals after controlling for the effects of observed covariates, which can be explained in terms of random effects for policyholders A and B.

[1] Most part of this chapter is from Jeong (2020).

Table 2.1: Hypothetical information on policyholders A and B

| Year | Gender | Age | Vehicle Size | Policyholder A | | Policyholder B | |
|---|---|---|---|---|---|---|---|
| | | | | # of Claim(s) | Claim Amt | # of Claim(s) | Claim Amt |
| 2015 | M | 45 | Medium | 0 | 0 | 1 | 500 |
| 2016 | M | 46 | Medium | 0 | 0 | 2 | 4000 |
| 2017 | M | 47 | Large | 1 | 200 | 1 | 8000 |

Because of the longitudinal property in most of P&C claim datasets, there have been some trials on the use of random effects model in actuarial science literature, which has a natural Bayesian interpretation. For example, Frangos and Vrontos (2001) tried to incorporate the random effects in bonus-malus system for automobile insurance and obtained a closed form formula for credibility premiums on compound loss, assuming the independence between the frequency and severity component. As an extension of their work, recently Jeong et al. (2020) also explored a random effects model for auto insurance claims considering possible dependence between the frequency and severity components.

Although the presence of random effects in the hypothetical example is very clear, it can be less clear in real longitudinal datasets observed by an insurance company. Therefore, one should be careful to incorporate random effects in a ratemaking model because it may capture random noise as unobserved heterogeneity via random effects so that the model has unnecessary complexity. However, a theoretical approach has not been attempted for testing the presence of random effects in claim modeling.

Intuitively, assuming absence of random effects on the heterogeneity of risk profiles for policyholders is equivalent to set the multiplicative random effects for all policyholders as a constant, which means the use of a point mass prior for random effects. Therefore, one can see that it is possible to test the presence of random effects in a longitudinal dataset via prior elicitation in Bayesian statistics. Bayesian inference requires to have an assumed prior distribution, which represents any a priori beliefs or uncertainties about the parameters. According to Dubitzky et al. (2013), "Elicitation is the process of extracting knowledge, beliefs, and uncertainties about unknown quantities from the client so that these can be expressed as a prior probability distribution."

In that sense, if a point mass prior is believed to be appropriate (or we have a strong belief that there is no presence of unobserved heterogeneity in the risks of policyholders) then it tells us that we can ignore the random effects in the modeling, whereas if a continuous prior is deemed to be suitable (or we have a strong belief that the impact of unobserved heterogeneity in the risks of policyholders is significant) then random effects are incorporated naturally in claim modeling.

It is possible to see some works on prior elicitation and Bayesian sensitivity analysis in actuarial literature though there is no previous direct work on testing the presence of random effects. For example, Gómez-Déniz et al. (1999) and Gómez-Déniz et al. (2000) performed Bayesian sensitivity analysis on Poisson-gamma frequency model to investigate how sensitive the posterior distribution of interest is to changes in prior distribution based on Esscher premium principle and variance premium principle, respectively.

In this chapter, Bregman divergence, proposed by Goh and Dey (2014) is used as a Bayesian model diagnostics for testing the robustness of a chosen prior. Since it is hardly possible to know the true prior distribution in general, we want the posterior distribution based on a chosen prior would not deviate too much from the true posterior distribution regardless of the true prior. Therefore, if the posterior distribution based on a continuous prior shows less deviation from the true posterior distribution compared to the posterior distribution based on a point mass prior, then we can favor a continuous prior as 'the more robust prior' and incorporate random effects in our model accordingly. This idea is applied to actuarial science so that we can test the presence of random effects in a longitudinal claim dataset and suggest a sophisticated framework for ratemaking model selection.

This chapter has been organized as follows. In Section 2.1, the two-part compound risk model is introduced and the models to be tested upon the presence of random effects are specified. In Section 2.2, the concept of Bregman divergence is introduced as well as the interpretation of Bregman divergence as a diagnostic for robustness of a chosen prior. In Section 2.3, description of the characteristics of the dataset and results of Bayesian sensitivity analysis are provided, which support the use of a continuous prior on the random effects rather than the use of a point mass. A conclusion is made in Section 2.4.

## 2.1 Longitudinal two-part compound risk model

Suppose that we have available information on $M$ policyholders for $T$ years. Then we can define the number of claims for the policyholder $i \in \{1, 2, \ldots, M\}$ in year $t$ as $N_{it}$. Likewise, the claim amount of the $k^{th}$ accident (where $k \leq N_{it}$) for the policyholder $i$ in year $t$ can be defined as $Y_{itk}$. Furthermore, we can define the exposure $e_{it} \in [0, 1]$ which means the proportion of coverage period within the calendar year $t$ for the policyholder $i$. Finally, we may define covariates $\mathbf{x}_{it}$, which often include age, gender, vehicle type, building type, building location, driving history and so forth. Note that each policyholder is followed up to $T_i \leq T$. Here $T_i$ means the number of insurance years for a specific policyholder $i$. Since it is not unusual for a policyholder to switch his/her insurance company once the automobile insurance contract expires, it is possible that $T_i$ varies for each policyholder.

For ratemaking in P&C insurance, it is of our interest to predict the following total cost of claims for each policyholder $i$ in year $t$:

$$S_{it} = \begin{cases} \sum_{k=1}^{N_{it}} Y_{itk}, & N_{it} \neq 0 \\ 0, & N_{it} = 0. \end{cases}$$

Then one can use two-part model to predict the number of claims $N_{it}$ and the average claim amount $C_{it}$ with the following decomposition of the joint density into the frequency part and conditional severity part:

$$f(N_{it}, C_{it}|\mathbf{x}_{it}) = f(N_{it}|\mathbf{x}_{it}) \times f(C_{it}|N_{it}, \mathbf{x}_{it}).$$

Note that $C_{it}$ is defined as

$$C_{it} = \begin{cases} \frac{1}{N_{it}} \sum_{k=1}^{N_{it}} Y_{itk}, & N_{it} \neq 0 \\ \text{Undefined}, & N_{it} = 0. \end{cases}$$

Here "Undefined" can be understood as "NA" because without observing any accident (in order words, $N = 0$), there is no way to observe the average claim amount per claim.

### 2.1.1 Frequency part model

In actuarial practices, Poisson distribution has been used for the calibration of the number of claims with the presence of covariates as follows:

$$N_{it}|\mathbf{x}_{it}, e_{it} \overset{indep}{\sim} \mathcal{P}(\nu_{it}) \quad \text{where} \quad \nu_{it} = e_{it} \exp(\mathbf{x}_{it}\alpha), \tag{2.1}$$

which means $N_{it}$ follows a Poisson distribution with mean $\nu_{it}$ and is independent to $N_{i't'}$ as long as $i \neq i'$ or $t \neq t'$ given the information on covariates and exposure.

Note that conditioning argument on $\mathbf{x}_{it}, e_{it}$ is suppressed afterward for notational convenience. Although this approach has been widely used due to its simplicity, the longitudinal property of usual claim datasets allows us to consider the unobserved heterogeneity of the policyholders via random effects as follows:

$$N_{it}|\theta_{N[i]} \overset{indep}{\sim} \mathcal{P}(\nu_{it}\theta_{N[i]}) \quad \text{where} \quad \nu_{it} = e_{it} \exp(\mathbf{x}_{it}\alpha), \;\; \theta_{N[i]} \sim \pi_N(\theta), \tag{2.2}$$

which has been explored by some authors, such as Boucher et al. (2008).

We can see that this random effects approach has a good Bayesian interpretation, because $\theta_{N[i]}$ is not observable so that we need to assume a prior on this. Furthermore, (2.1) can be interpreted as a special case of (2.2) with $\mathbb{P}(\theta_{N[i]} = \theta) = \mathbb{1}_{\{\theta=1\}}$ for all $i$ where $\mathbb{1}_{\{\theta=1\}} = 1$ if

$\theta = 1$ and $\mathbb{1}_{\{\theta=1\}} = 0$ otherwise.

Therefore, model selection between (2.1) and (2.2) is equivalent to the prior elicitation of $\pi_N(\theta)$. Since the impact of unobserved heterogeneity on claim frequency is usually unknown so that it needs to be assessed by the observed claim frequency, it is desirable to use noninformative prior on $\theta_N$, which has less impact on our Bayesian analysis. Therefore, as in a lot of Bayesian literature including but not limited to Jeffreys (1946) and Berger (1985), it is natural to consider the use of the Jeffreys' prior as a candidate of noninformative prior on $\theta_N$. Note that the Jeffreys' prior for a random variable $Z$ with density $f(z|\theta)$ is defined as the square root of the Fisher information $I(\theta)$, where $I(\theta) = \mathbb{E}\left[-\frac{\partial^2}{\partial\theta^2}\log f(z|\theta)\right]$.

Suppose $N|\theta_N \sim \mathcal{P}(\nu\theta_N)$ where $\theta_N > 0$. Then the Jeffreys' prior of $\theta_N$ is given as $\pi_N(\theta) = \theta^{-1/2}$ and the corresponding posterior distribution is gamma distribution with the following density:

$$\pi_N(\theta|N) \propto \theta^{N-1/2}\exp(-\nu\theta) \;\text{ so that }\; \theta_N|N \sim \mathcal{G}(N+0.5,\, \nu^{-1})$$

because it is easy to see that

$$I(\theta) = \mathbb{E}\left[-\frac{\partial^2}{\partial\theta^2}\log p(N|\theta)\right] = \nu/\theta \implies \pi_N(\theta) = \theta^{-1/2} \propto \sqrt{I(\theta)}$$

and

$$\pi_N(\theta|N) \propto \pi_N(\theta)p(N|\theta) \propto \theta^{N-1/2}\exp(-\nu\theta) \implies \theta_N|N \sim \mathcal{G}(N+0.5,\, \nu^{-1}).$$

Note that although the Jeffreys' prior is improper, the corresponding posterior is still proper. However, even though we can have a prior with less information and proper posterior, we hope the mean of $\theta_N$ to be one because $\theta_N$ is a multiplicative random effect and it gives rise to an identifiability issue if $\mathbb{E}[\theta_N] \neq 1$. Therefore, we want to impose $\mathbb{E}[\theta_N] = 1$ as shown in Ng and Cook (2000) and Ding and Wang (2008) while we retain the same distribution on the posterior. Hence, we may propose the following prior on $\theta_N$, which satisfies $\mathbb{E}\left[\theta_N\right] = 1$ as well as has conjugate gamma posterior.

$$\pi_N(\theta) \propto \theta^{r-1}e^{-\theta r} \text{ so that } \theta_{N[i]} \overset{i.i.d.}{\sim} \mathcal{G}(r, 1/r),\ \mathbb{E}\left[\theta_{N[i]}\right] = 1, \text{ and } \mathbb{V}ar\left[\theta_{N[i]}\right] = \frac{1}{r}. \tag{2.3}$$

Now, according to the aforementioned arguments, we can formulate the model selection in case of frequency part as follows:

[Naive Frequency Model]

Data likelihood: $N_{it}|\theta_{N[i]} \overset{indep}{\sim} \mathcal{P}(\nu_{it}\theta_{N[i]})$

Prior distribution: $\mathbb{P}(\theta_i^N = \theta) = \mathbb{1}_{\{\theta=1\}}$ for all $i$.

Posterior distribution:

$$\mathbb{P}(\theta_i^N = \theta | N_{i1} = n_{i1}, \ldots, N_{iT_i} = n_{iT_i}) \propto \mathbb{P}(\{\theta_i^N = \theta\} \cap \{N_{i1} = n_{i1}, \ldots, N_{iT_i} = n_{iT_i}\}) \propto \mathbb{1}_{\{\theta=1\}}.$$

Predictive distribution:

$$p(N_{i,T_i+1}|N_{i1}, N_{i2}, \ldots, N_{iT_i}) = \int p(N_{i,T_i+1}|\theta)\pi_N(\theta|N_{i1}, N_{i2}, \ldots, N_{iT_i})d\theta = p(N_{i,T_i+1}|\theta = 1).$$

Therefore, we can see that $N_{i,T_i+1}|N_{i1}, N_{i2}, \ldots, N_{iT_i} \sim \mathcal{P}(\nu_{i,T_i+1})$, which means that predictive distribution of $N_{i,T_i+1}$ does not depend on the previous claim frequency observation due to the marginal independence among $N_{i,t}$.

[Proposed Frequency Model]

Data likelihood: $N_{it}|\theta_{N[i]} \overset{indep}{\sim} \mathcal{P}(\nu_{it}\theta_{N[i]})$

Prior distribution: $\pi_N(\theta) \propto \theta^{r-1}e^{-\theta r}$ so that $\theta_{N[i]} \sim \mathcal{G}(r, 1/r)$ and $\mathbb{E}\left[\theta_{N[i]}\right] = 1, \mathbb{V}ar\left[\theta_{N[i]}\right] =$

$\frac{1}{r}$. Therefore, as $r \to \infty$, $\pi_N(\theta)$ degenerates to the Dirac delta function at $\theta = 1$ which means that the naive frequency model is a merely limiting case of the proposed frequency model. According to Lemaire (1998), the observed number of claims from previous years has been widely used as an adjustment weight factor to penalize or provide bonus on policyholders, which is analogous to the empirical estimates of the values of random effects on claim frequency. Moreover, the range of adjustment weight factor on frequency premiums is usually from 54% to 200%. Therefore, it is natural to incorporate this knowledge on choosing the hyperparameter $r$ for our proposed prior so that the 95% highest posterior density (HPD) interval of $\theta_N$ can include $(0.54, 2.00)$. Thus, $r = 3.8$ is used as the hyperparameter so that 95% HPD interval of $\theta_N$ under the proposed prior can be around $(0.16, 2.01)$.

Posterior distribution:

$$
\begin{aligned}
\pi_N(\theta|N_{i1}, N_{i2}, \ldots, N_{iT_i}) &\propto \pi_N(\theta) \prod_{t=1}^{T_i} p(N_{it}|\theta) \propto \theta^{r-1} e^{-\theta r} \left( \prod_{t=1}^{T_i} \theta^{N_{it}} \right) e^{-\sum_{t=1}^{T_i} \nu_{it}\theta} \\
&\propto \theta^{\sum_{t=1}^{T_i} N_{it}+r-1} e^{-\theta(\sum_{t=1}^{T_i} \nu_{it}+r)}
\end{aligned}
$$

so that $\theta_{N[i]}|N_{i1}, N_{i2}, \ldots, N_{iT_i} \sim \mathcal{G}(\sum_{t=1}^{T_i} N_{it} + r, [\sum_{t=1}^{T_i} \nu_{it} + r]^{-1})$.

Predictive distribution:

$$
\begin{aligned}
p(N_{i,T_i+1}|N_{i1}, N_{i2}, \ldots, N_{iT_i}) &= \int p(N_{i,T_i+1}|\theta)\pi_N(\theta|N_{i1}, N_{i2}, \ldots, N_{iT_i})d\theta \\
&= \binom{\sum_{t=1}^{T_i+1} N_{it} + r - 1}{N_{i,T_i+1}} \left( \frac{\sum_{t=1}^{T_i} \nu_{it} + r}{\sum_{t=1}^{T_i+1} \nu_{it} + r} \right)^{\sum_{t=1}^{T_i} N_{it}+r} \left( \frac{\nu_{i,T_i+1}}{\sum_{t=1}^{T_i+1} \nu_{it} + r} \right)^{N_{i,T_i+1}}
\end{aligned}
$$

Therefore, we can see that

$$
N_{i,T_i+1}|N_{i1}, N_{i2}, \ldots, N_{iT_i} \sim \mathcal{NB}\left( \sum_{t=1}^{T_i} N_{it} + r, \frac{\nu_{i,T_i+1}}{\sum_{t=1}^{T_i+1} \nu_{it} + r} \right)
$$

so that $\mathbb{E}\left[N_{i,T_i+1}|N_{i1}, N_{i2}, \ldots, N_{iT_i}\right] = \frac{\sum_{t=1}^{T_i} N_{it}+r}{\sum_{t=1}^{T_i} \nu_{it}+r} \nu_{i,T_i+1}$.

Note that individual premium on the frequency component depends on random effect $\theta_N$ as well as the covariate information at time $t$, which is associated with the regression coefficient $\alpha$ so that we have the following:

$$\mathbb{E}[N_{i,T_i+1}|N_{i,1},\dots,N_{iT_i}] = \exp(\mathbf{x}_{i,T_i+1}\alpha)\,\mathbb{E}[\theta_{N[i]}|N_{i1},\dots,N_{iT_i}],$$

which means posterior mean of $\theta_{N[i]}$ is not the same as the predictive mean of $N_{i,T_i+1}$ given $N_{i1},\dots,N_{iT}$. Furthermore, knowing predictive distribution of $N_{i,T_i+1}|N_{i1},\dots,N_{iT}$ could be useful since $\mathbb{E}[N_{i,T_i+1}e^{\gamma N_{i,T_i+1}}|N_{i1},\dots,N_{iT}]$ needs to be evaluated in order to obtain $\mathbb{E}[S_{i,T_i+1}|\,N_{i1},\dots,N_{iT_i},C_{i1},\dots,C_{iT_i}]$ with dependence between the frequency and severity components.

### 2.1.2 Severity part model

Traditionally, gamma distribution has been used for the calibration of the average claim amount with the presence of covariates as follows:

$$C_{it}|\mathbf{x}_{it},N_{it} \overset{indep}{\sim} \mathcal{G}(\psi_{it},\mu_{it}/\psi_{it}) \quad \text{where} \;\; \mu_{it}=\exp(\mathbf{x}_{it}\beta+\gamma N_{it}),\;\; \psi_{it}=N_{it}/\phi. \tag{2.4}$$

Note that conditioning argument on $\mathbf{x}_{it}, N_{it}$ is suppressed afterward for notational convenience. Again, the longitudinal property of usual claim datasets allows us to consider the unobserved heterogeneity of the policyholders via random effects as follows:

$$C_{it}|\theta_{C[i]} \overset{indep}{\sim} \mathcal{G}(\psi_{it},\theta_{C[i]}\mu_{it}/\psi_{it}) \quad \text{where} \;\; \mu_{it}=\exp(\mathbf{x}_{it}\beta+\gamma N_{it}),\;\; \psi_{it}=N_{it}/\phi,\;\; \theta_{C[i]}\sim\pi_C(\theta). \tag{2.5}$$

Unlike the traditional approach for compound loss model which assumes independence between the frequency and severity components, here the observed frequency is also used as an explanatory variable for the average severity to capture the possible dependence between the frequency and the average severity. Although the independence assumption between the frequency and severity has been widely used due to its simplicity, recent research works in actuarial science show empirical evidences of dependence between the frequency and severity in various claim datasets. For the detailed approach, please see Garrido et al. (2016) and Jeong et al. (2020). We also have a good Bayesian interpretation in this case so that (2.4) can be interpreted as a special case of (2.5) with $\mathbb{P}(\theta_i^C = \theta) = \mathbb{1}_{\{\theta=1\}}$ for all $i$.

Therefore, model selection between (2.4) and (2.5) is equivalent to the prior elicitation of $\pi_C(\theta)$ and we again consider the use of the Jeffreys' prior as a candidate of noninformative prior on $\theta_C$ as follows. Suppose $C|\theta_C \sim \mathcal{G}(\psi, \mu\theta_C/\psi)$ where $\theta_C > 0$. Then the Jeffreys' prior of $\theta_C$ is given as $\pi_C(\theta) = \theta^{-1}$ and the corresponding posterior distribution is inverse gamma distribution with the following density:

$$\pi_C(\theta|C) \propto \left(\frac{1}{\theta}\right)^{-\psi-1} \exp\left(-\frac{\psi C\mu^{-1}}{\theta}\right) \quad \text{so that} \quad \theta_C|C \sim \mathcal{IG}(\psi,\, \psi C\mu^{-1})$$

since it is easy to see that

$$I(\theta) = \mathbb{E}\left[-\frac{\partial^2}{\partial\theta^2}\log f(C|\theta)\right] = \psi/\theta^2 \implies \pi_C(\theta) = \theta^{-1} \propto \sqrt{I(\theta)}$$

and

$$\pi_C(\theta|C) \propto \pi_C(\theta) f(C|\theta) \propto \left(\frac{1}{\theta}\right)^{-\psi-1} \exp\left(-\frac{\psi C\mu^{-1}}{\theta}\right) \implies \theta_C|C \sim \mathcal{IG}(\psi,\, \psi C\mu^{-1}).$$

Note that although the Jeffreys' prior is improper, the corresponding posterior is still proper. However, again we hope the mean of $\theta_C$ to be one because $\theta_C$ is a multiplicative random effect. Hence, we may propose the following prior on $\theta_C$, which satisfies $\mathbb{E}[\theta_C] = 1$ as well as has conjugate inverse gamma posterior.

$$\pi_C(\theta) \propto \theta^{-k-2}e^{-k/\theta} \text{ so that } \theta_{C[i]} \overset{i.i.d.}{\sim} \mathcal{IG}(k+1,k),\ \mathbb{E}\left[\theta_{C[i]}\right]=1, \text{ and } \mathrm{V}ar\left[\theta_{C[i]}\right]=\frac{1}{k-1}. \tag{2.6}$$

Therefore, we can formulate the model selection in case of the average severity part as follows:

[Naive Severity Model]

Data likelihood: $C_{it}|\theta_{C[i]} \overset{indep}{\sim} \mathcal{G}(\psi_{it}, \theta_{C[i]}\mu_{it}/\psi_{it})$

Prior distribution: $\mathbb{P}(\theta_i^C=\theta)=\mathbb{1}_{\{\theta=1\}}$ for all $i$.

Posterior distribution:

$$\mathbb{P}(\theta_i^C=\theta|C_{i1}=c_{i1},\ldots,C_{iT_i}=c_{iT_i}) \propto \mathbb{P}(\{\theta_i^C=\theta\}\cap\{C_{i1}=c_{i1},\ldots,C_{iT_i}=c_{iT_i}\}) \propto \mathbb{1}_{\{\theta=1\}}.$$

Predictive distribution:

$$\begin{aligned} f(C_{i,T_i+1}|C_{i1},C_{i2},\ldots,C_{iT_i}) &= \int f(C_{i,T_i+1}|\theta)\pi_C(\theta|C_{i1},C_{i2},\ldots,C_{iT_i})d\theta \\ &= f(C_{i,T_i+1}|\theta=1). \end{aligned}$$

Therefore, we can see that $C_{i,T_i+1}|C_{i1},C_{i2},\ldots,C_{iT_i} \sim \mathcal{G}(\psi_{i,T_i+1}, \mu_{i,T_i+1}/\psi_{i,T_i+1})$, which means that predictive distribution of $C_{i,T_i+1}$ does not depend on the previous claim severity observations due to the marginal independence among $C_{i,t}$.

[Proposed Severity Model]

Data likelihood: $C_{it}|\theta_{C[i]} \overset{indep}{\sim} \mathcal{G}(\psi_{it}, \theta_{C[i]}\mu_{it}/\psi_{it})$

Prior distribution: $\pi_C(\theta) \propto \theta^{-k-2}e^{-k/\theta}$ so that $\theta_{C[i]} \sim \mathcal{IG}(k+1,k)$ and $\mathbb{E}\left[\theta_{C[i]}\right]=1, \mathrm{V}ar\left[\theta_{C[i]}\right]=\frac{1}{k-1}$. Therefore, as $k\to\infty$, $\pi_C(\theta)$ degenerates to the Dirac delta function at

$\theta = 1$, which means that the naive severity model is a merely limiting case of the proposed severity model. According to the Lemaire (1998), most of countries except for South Korea do not use historically observed claim amounts for the construction of penalty or bonus on a policyholder, which supports the assertion that there is less variability on $\theta_C$, the random effect of the severity component than on $\theta_N$, the random effect of the frequency component. Therefore, $k = 11$ is used so that the 95% HPD interval of $\theta_C$ under the proposed prior can be around $(0.49, 1.61)$, which is narrower than the 95% HPD interval of $\theta_N$ under the proposed prior. Indeed, if Empirical Bayes method is applied by maximizing the marginal likelihood with respect to both $\beta$ and $k$ where initial value of $k$ as 11, then the optimal $k$ is given as 11.00226. However, since five digits of decimal might give a false feeling of precision and 11.00226 is not much different from 11, $k = 11$ is used as the value of hyperparameter throughout this chapter.

Posterior distribution:

$$\begin{aligned}\pi_C(\theta|C_{i1}, C_{i2}, \ldots, C_{iT_i}) &\propto \pi_C(\theta) \prod_{t=1}^{T_i} f(C_{it}|\theta) \propto \theta^{-k-2} e^{-k/\theta} \left( \prod_{t=1}^{T_i} \theta^{-\psi_{it}} \right) e^{-(\sum_{t=1}^{T_i} \frac{\psi_{it} C_{it}}{\mu_{it}})/\theta} \\ &\propto \theta^{-(\sum_{t=1}^{T_i} \psi_{it} + k + 2)} e^{-(\sum_{t=1}^{T_i} \frac{\psi_{it} C_{it}}{\mu_{it}} + k)/\theta}\end{aligned}$$

so that $\theta_{C[i]}|C_{i1}, C_{i2}, \ldots, C_{iT_i} \sim \mathcal{IG}(k + \sum_{t=1}^{T_i} \psi_{it} + 1, \sum_{t=1}^{T_i} \frac{\psi_{it} C_{it}}{\mu_{it}} + k)$.

Predictive distribution:

$$\begin{aligned}f(C_{i,T_i+1}|C_{i1}, C_{i2}, \ldots, C_{iT_i}) &= \int f(C_{i,T_i+1}|\theta) \pi_C(\theta|C_{i1}, C_{i2}, \ldots, C_{iT_i}) d\theta \\ &= \frac{(\psi_{i,T_i+1} C_{i,T_i+1}/\mu_{i,T_i+1})^{\psi_{i,T_i+1}}}{(k + \sum_{t=1}^{T_i+1} \psi_{i,t} C_{i,t}/\mu_{i,t})^{\sum_{t=1}^{T_i} \psi_{i,t} + k + 1}} \times \frac{\Gamma(\sum_{t=1}^{T_i} \psi_{i,t} + k + 1) C_{i,T_i+1}^{-1}}{\Gamma(\psi_{i,T_i+1}) \Gamma(\sum_{t=1}^{T_i} \psi_{i,t} + k + 1)}\end{aligned}$$

Therefore, we can see that

$$C_{i,T_i+1}|C_{i1}, C_{i2}, \ldots, C_{iT_i} \sim \mathcal{GP}\left( k + \sum_{t=1}^{T_i} \psi_{it} + 1, \left[ k + \sum_{t=1}^{T_i} \psi_{i,t} \frac{C_{it}}{\mu_{it}} \right] \frac{\mu_{i,T_i+1}}{\psi_{i,T_i+1}}, \psi_{i,T_i+1} \right),$$

with predictive mean

$$\mathbb{E}\left[C_{i,T_i+1}|C_{i1}, C_{i2}, \ldots, C_{iT_i}\right] = \frac{\left[k + \sum_{t=1}^{T_i} \psi_{it}\frac{C_{it}}{\mu_{it}}\right]}{k + \sum_{t=1}^{T_i} \psi_{it}}\mu_{i,T_i+1} = \frac{\left[k\phi + \sum_{t=1}^{T_i} S_{it}/\mu_{it}\right]}{k\phi + \sum_{t=1}^{T_i} N_{it}}\mu_{i,T_i+1}$$

since $\psi_{it}C_{it} = N_{it}C_{it}/\phi = S_{it}/\phi$.

Note that generalized Pareto (GP) distribution is defined with the following density as in Klugman et al. (2012):

$$f(y|a, \tau, c) = \frac{\Gamma(a+\tau)}{\Gamma(a)\Gamma(\tau)}\frac{c^a x^{\tau-1}}{(x+c)^{a+\tau}}$$

so that $\mathbb{E}\left[Y\right] = c\frac{\tau}{a-1}$ when $Y \sim \mathcal{GP}(a, \tau, c)$.

One can see that the suggested compound model leaves the patterns that are usually used in the actuarial field, the underlying assumption of independence between the frequency and severity components by letting $\mathbb{E}[C_{it}|N_{it}, \theta_{C[i]}] = \theta_{C[i]}\mu_{it} = \theta_{C[i]}\exp(\mathbf{x}_{it}\beta + N_{it}\gamma)$, which is a flexible extension of traditional independent compound risk model so that we may capture possible dependence between the frequency and severity component via $\gamma$. It is an interesting topic by itself to understand a posteriori premium of $S_{i,T_i+1}$ with both types of dependences, dependence between frequency and severity as well as dependence among the claims of the same policyholder. However, a thorough discussion on this topic is not covered in this chapter in order not to overwhelm the readers.

## 2.2 Bayesian sensitivity analysis with Bregman divergence

Bayesian sensitivity analysis, which is also known as robust Bayesian analysis, is an area which studies the impact on the posterior due to possible perturbations of prior distribution. If there is less impact on the posterior even after perturbations of given prior, then we can claim that given prior is relatively robust to perturbations so that there is more consistency on the Bayesian analysis, which is based on the obtained posterior distribution of parameter(s). There are some proposed methods for modeling perturbations of prior distribution which are described in Berger et al. (1994), but here we can use the comparison between given $\pi(\theta)$ and the $\epsilon$-contaminated prior $\pi_\epsilon(\theta)$, which is defined as follows:

$$\Pi = \{\pi_\epsilon(\theta) : \pi_\epsilon(\theta) = (1-\epsilon)\pi(\theta) + \epsilon\, q(\theta), q \in \mathbf{Q}, \epsilon \in [0,1]\}, \tag{2.7}$$

where $\mathbf{Q}$ is a certain class of prior distributions. Note that $\epsilon$-contaminated prior has been used in the actuarial literature such as Gómez-Déniz et al. (2002b), Gómez-Déniz and Vázquez-Polo (2005), and Gómez-Déniz et al. (2006).

Here $\pi_\epsilon(\theta)$ is a prior distribution which can capture possible perturbations of the prior so that high value of $\epsilon$ means that the $\pi_\epsilon(\theta)$, which is assumed to be the true prior, could be far from $\pi(\theta)$, the proposed prior. Therefore, it is our interest to minimize the 'distance' between $\pi_\epsilon(\theta|z)$ and $\pi(\theta|z)$ by choosing appropriate $\pi(\theta)$ regardless of possible perturbations of the true prior.

To measure the distance between the posterior densities, we can consider the concept of Bregman divergence proposed by Bregman (1967). Bregman divergence has some properties of usual metric but neither is symmetric nor satisfies the triangle inequality. After the introduction of its concept, it has been utilized in a variety of statistical learning.

For example, Gelfand and Dey (1991) used Kullback-Leibler divergence in Bayesian sensitivity analysis while Peng and Dey (1995) applied $f$-divergence in the context of outlier detection. Note that both KL divergence and $f$-divergence can be explained in terms of Bregman divergence. Zhang (2004) used Bregman divergence to study statistical behavior and consistency of classification methods. Recently, Bregman divergence was used to obtain a class of loss function in order for robust Bayesian prediction according to Karimnezhad and Parsian (2018).

According to Goh and Dey (2014), the difference between $\pi_\epsilon(\theta|z)$ and $\pi(\theta|z)$ can be measured by using functional Bregman divergence which is defined as follows:

**Definition 2.1.** *Let $h_1, h_2$ be non-negative measurable functions on $\sigma-$finite measure space $(\mathcal{Z}, \Omega, v)$ and $\psi : (0, \infty) \to \mathbb{R}$ be a strictly convex and differentiable function. Then the functional Bregman divergence $D_\psi$ is defined as*

$$D(h_1, h_2) = \int \{\psi(h_1(z)) - \psi(h_2(z)) - (h_1(z) - h_2(z))\psi'(h_2(z))\} \, dv(z).$$

It is easy to check that $D(h, h) = 0$ for any non-negative measurable function $h$. Therefore, it is desirable to minimize the following quantity, which measures the relative difference between $\pi_\epsilon(\theta|z)$ and $\pi(\theta|z)$:

$$\begin{aligned}
D_\psi^R = D_\psi\left(\frac{\pi_\epsilon(\theta|z)}{\pi(\theta|z)}, 1\right) &= \int \left\{\psi\left(\frac{\pi_\epsilon(\theta|z)}{\pi(\theta|z)}\right) - \psi(1) - \left(\frac{\pi_\epsilon(\theta|z)}{\pi(\theta|z)} - 1\right)\psi'(1)\right\} \pi(\theta|z)d\theta \\
&= \int \left\{\psi\left(\frac{\pi_\epsilon(\theta|z)}{\pi(\theta|z)}\right)\pi(\theta|z) - \psi(1)\pi(\theta|z) - (\pi_\epsilon(\theta|z) - \pi(\theta|z))\,\psi'(1)\right\} d\theta \qquad (2.8) \\
&= \int \psi\left(\frac{\pi_\epsilon(\theta|z)}{\pi(\theta|z)}\right)\pi(\theta|z)d\theta - \psi(1)
\end{aligned}$$

Note that if we can obtain the closed forms of both $\pi_\epsilon(\theta)$ and $\pi(\theta)$, and furthermore it is easy to evaluate the integral given in the end of (2.8), then it might be okay to directly use the closed form as in the following lemma.

**Lemma 2.1.** *Suppose $f(z|\theta)$ is data likelihood and $\mathbb{P}(\theta = 1) = 1$, in other words, $\theta$ has the* *point mass at 1. Then $D_\psi^R$ is given as follows:*

$$
\begin{aligned}
D_\psi^R &= \psi\left(\frac{\pi_\epsilon(1|z)}{\pi(1|z)}\right) - \psi(1) \\
&= \psi\left(\frac{\pi_\epsilon(1)f(z|1)}{m_\epsilon(1|z)}\frac{1}{\pi(1|z)}\right) - \psi(1) \\
&= \psi\left(\frac{(1-\epsilon)\pi(1) + \epsilon\, q(1)}{(1-\epsilon)m(z) + \epsilon\, m_q(z)}\frac{f(z|1)}{\pi(1|z)}\right) - \psi(1) \\
&= \psi\left(\frac{(1-\epsilon) + \epsilon\, q(1)/\pi(1)}{(1-\epsilon) + \epsilon\, m_q(z)/f(z|1)}\right) - \psi(1),
\end{aligned}
$$

*where $m(z) = \int f(z|\theta)\pi(\theta)d\theta$ and $m_q(z) = \int f(z|\theta)q(\theta)d\theta$.*

*Proof.* Since $\theta$ has the point mass at 1, we have

$$\int \psi\left(\frac{\pi_\epsilon(\theta|z)}{\pi(\theta|z)}\right)\pi(\theta|z)d\theta = \psi\left(\frac{\pi_\epsilon(1|z)}{\pi(1|z)}\right).$$

Furthermore, since $\mathbb{P}(\theta = 1|z) = 1$ as well, $\pi(\theta)$ and $\pi(\theta|z)$ are the same as the Dirac delta function at $\theta = 1$. Therefore,

$$m(z) = \int f(z|\theta)\pi(\theta)d\theta = f(z|1),\ \pi(1) = \pi(1|z).$$

□

Since $\pi(\theta)$ is the Dirac delta function, it has infinite value when $\theta = 1$. Therefore, in actual implementation, we may use $\tilde{\pi}(\theta) \sim \mathcal{N}(1, 10^{-12})$.

However, in most of cases, it might not be possible to obtain the closed form of either $\pi_\epsilon(\theta)$ or $\pi(\theta)$. Even though we have the closed forms, still we are not sure whether the integral is able to be evaluated in an analytic way. Therefore, by denoting $\delta(\theta) := \pi_\epsilon(\theta)/\pi(\theta)$, we may use the following equation as shown in Goh and Dey (2014), which is equivalent to (2.8) but enables us to implement MCMC algorithm to evaluate $D_{\psi}^{R}$ numerically.

$$\begin{aligned}
D_{\psi}^{R}+\psi(1) &= \int \psi\left(\frac{\pi_{\epsilon}(\theta|z)}{\pi(\theta|z)}\right)\pi(\theta|z)d\theta \\
&= \int \psi\left(\frac{\pi_{\epsilon}(\theta)f(z|\theta)}{\pi(\theta|z)\int \pi_{\epsilon}(\theta)f(z|\theta)d\theta}\right)\pi(\theta|z)d\theta = \int \psi\left(\frac{\delta(\theta)\pi(\theta)f(z|\theta)}{\pi(\theta|z)\int \delta(\theta)\pi(\theta)f(z|\theta)d\theta}\right)\pi(\theta|z)d\theta \\
&= \int \psi\left(\frac{\delta(\theta)\pi(\theta|z)}{\pi(\theta|z)\int \delta(\theta)\pi(\theta|z)d\theta}\right)\pi(\theta|z)d\theta = \int \psi\left(\frac{\delta(\theta)}{\int \delta(\theta)\pi(\theta|z)d\theta}\right)\pi(\theta|z)d\theta \\
&\simeq \frac{1}{J}\sum_{j=1}^{J}\left[\psi\left(\frac{\delta(\hat{\theta}_j)}{\frac{1}{J}\sum_{j=1}^{J}\delta(\hat{\theta}_j)}\right)\right],
\end{aligned} \tag{2.9}$$

where $\hat{\theta}_j$ s are posterior samples derived from $\pi(\theta|z)$.

Finally, to perform the sensitivity analysis with the contaminated class of priors in (2.7), it is desirable to choose $\mathbf{Q}$ carefully so that it might neither be too broad nor narrow, as mentioned in Berger et al. (1986). Therefore, here we can consider the family $\mathbf{Q}$ which satisfies the usual assumption of multiplicative random effects, having 1 as the prior mean. Furthermore, since we are not sure whether the naive prior or the proposed prior represents true dynamics on $\theta$, we consider the family of distribution $\mathbf{Q}$ which has the average of standard deviations of $\theta$ under the naive and proposed priors as the standard deviation of $\theta$ with $q(\theta)$. One can see that there are some research works which specified the class $\mathbf{Q}$ in terms of moments, including but not limited to Eichenauer et al. (1988), Young (1998), Insua et al. (1999), Gómez-Déniz et al. (2002a), Gómez-Déniz et al. (2005), Boratyńska (2017), and Sánchez-Sánchez et al. (2019).

Therefore, $\mathbf{Q}$ can be defined as follows:

$$\mathbf{Q} = \left\{ q(\theta) : \mathbb{E}^{q}[\theta] = 1. \quad \mathbb{V}ar^{q}[\theta] = \frac{1}{4}(\mathbb{V}ar^{p}[\theta] + \mathbb{V}ar^{n}[\theta]) = \frac{\mathbb{V}ar^{p}[\theta]}{4} \right\}.$$

Here $\mathbb{V}ar^{n}[\theta]$ means the variance of $\theta$ under the naive prior and $\mathbb{V}ar^{p}[\theta]$ means the variance of $\theta$ under the proposed prior. And it is also easy to see that $\mathbb{V}ar^{n}[\theta] = 0$ and $\sqrt{\mathbb{V}ar^{q}[\theta]} =$

$\sqrt{\mathbb{V}ar^p[\theta]}/2$.

Therefore, in the following section, we are considering uniform, lognormal, and normal priors as possible perturbations for $\theta_N$ and $\theta_C$, respectively so that they can satisfy both mean and variance constraints as follows:

$$\text{For } \theta_N : q_1(\theta) \sim \mathcal{U}(0.5557, 1.4443),\ q_2(\theta) \sim \mathcal{LN}(-0.0319, 0.2524),\ q_3(\theta) \sim \mathcal{N}(1, 0.0658),$$

$$\text{For } \theta_C : q_1(\theta) \sim \mathcal{U}(0.7261, 1.2738),\ q_2(\theta) \sim \mathcal{LN}(-0.0123, 0.1571),\ q_3(\theta) \sim \mathcal{N}(1, 0.0250).$$

## 2.3 Data analysis

For the empirical analysis, a public dataset on insurance claims provided by Wisconsin Local Government Property Insurance Fund (LGPIF) is used, which has been used in actuarial literature such as Frees et al. (2016b). It consists of 5,677 observations in traning set and 1,098 observations in test set. It is a longitudinal dataset with 1,234 policyholders which can be tracked with a unique identifier, followed for 5 years on multiple lines of claims. Among the information on multi-line insurance, only inland marine (IM) claim information was used. Given dataset has seven categorical explanatory variables, most of which are indicator variables on the types of location. Note that 'NoClaimCreditIM' is used in both frequency and severity modeling considering current practices in ratemaking, because a premium discount is followed by the absence of claim for three consecutive prior years, as a rule of thumb in practice.

As continuous variables, the coverage amount of IM claim and deductible amount for IM claim were used, which are expected to have positive and negative effects on the claims, respectively.

Table 2.2: Observable policy characteristics used as covariates

| Categorical variables | Description | | Proportions | |
|---|---|---|---|---|
| TypeCity | Indicator for city entity: | Y=1 | 14 % | |
| TypeCounty | Indicator for county entity: | Y=1 | 5.78 % | |
| TypeMisc | Indicator for miscellaneous entity: | Y=1 | 11.04 % | |
| TypeSchool | Indicator for school entity: | Y=1 | 28.17 % | |
| TypeTown | Indicator for town entity: | Y=1 | 17.28 % | |
| TypeVillage | Indicator for village entity: | Y=1 | 23.73 % | |
| NoClaimCreditIM | No IM claim in three consecutive prior years: | Y=1 | 42.1 % | |
| Continuous variables | | Minimum | Mean | Maximum |
| CoverageIM | Log coverage amount of IM claim in mm | 0 | 0.85 | 46.75 |
| lnDeductIM | Log deductible amount for IM claim | 0 | 5.34 | 9.21 |

In order to apply the idea of capturing individual heterogeneity via random effect, we should assume that "The same person or object" is followed for many years by a unique identifier even though the characteristics of insurance contract change, and the source of individual heterogeneity is consistent for observed years. For inland marine insurance data described above, one can see that it satisfies both assumptions since a specific object can be observed repeatedly via a unique classifier. However, in case of automobile insurance, validity of the assumptions could be controversial. For example, it is possible a policyholder shares a car with his/her kid or his/her driving skills (one of the unobserved risk characteristics) might be improved over time. A thorough discussion on models with varying or multiple sources of random effects could be an interesting topic for future research.

In terms of frequency, IM has relatively moderate dispersion of the number of claims per year so that maximum number of claims per year is six. Since the observed sample mean of the number of claims is much smaller than the observed sample variance, it is natural to consider the use of different types of frequency distribution on the modeling other than naive Poisson distribution. Moreover, it can be shown that marginal distribution of claim frequency follows a multivariate negative binomial (MVNB) distribution under the proposed prior so that it provides another rationale to consider a non-point mass prior on the random effect of the frequency component.

Table 2.3: Summary statistics for claim frequency

| | | Minimum | Mean | Variance | Maximum |
|---|---|---|---|---|---|
| FreqIM | number of IM claims in a year | 0 | 0.06 | 0.1 | 6 |

Table 2.4: Distribution of IM frequency

| Count | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| FreqIM | 5441 | 182 | 40 | 6 | 4 | 2 | 2 |

Table 2.5: Summary statistics for IM severity

| | | Minimum | Mean | Variance | Maximum |
|---|---|---|---|---|---|
| log(yAvgIM) | (log) avg size of IM claim in a year | 4.09 | 8.45 | 2.23 | 13.09 |

After we fixed the hyperparameters on the priors of each random effect component, the marginal likelihood of both frequency and the average severity components could be obtained with the naive and proposed priors, respectively. Upon the obtained marginal likelihood, as a type of empirical Bayes method, the regression coefficients can be estimated with the marginal likelihood and observed data. For the frequency component, use of the naive prior leads us to the marginal likelihood independent Poisson distribution, whereas the use of proposed prior leads us to the marginal likelihood of MVNB distribution. Under each marginal likelihood, $\hat{\alpha}$ were obtained, which are hyperparameters in the frequency model associated with the explanatory variables. The estimated values are shown in Table 2.6.

Table 2.6: Regression estimates from marginal frequency likelihoods

| | Poisson | | MVNB | |
|---|---|---|---|---|
| | Estimate | Std. Error | Estimate | Std. Error |
| (Intercept) | -6.9455 | 1.0211 | -7.3601 | 1.1532 |
| TypeCity | 3.7219 | 1.0101 | 3.7844 | 1.1359 |
| TypeCounty | 4.5654 | 1.0124 | 4.6135 | 1.1409 |
| TypeSchool | 1.8423 | 1.0274 | 2.0799 | 1.1496 |
| TypeTown | 2.3378 | 1.0263 | 2.5526 | 1.1494 |
| TypeVillage | 2.7545 | 1.0139 | 2.9450 | 1.1383 |
| CoverageIM | 0.0647 | 0.0072 | 0.0946 | 0.0143 |
| lnDeductIM | 0.1531 | 0.0455 | 0.1732 | 0.0520 |
| NoClaimCreditIM | -0.3697 | 0.1283 | -0.1985 | 0.1326 |

In case of the average severity component, use of the naive prior leads us to independent gamma marginal likelihood, whereas use of the proposed prior leads us to marginal likelihood of multivariate generalized Pareto (MVGP) distribution. Again, under each marginal likelihood, $\hat{\beta}$ and $\hat{\gamma}$ were obtained, which are hyperparameters in the average severity model associated with the explanatory variables. The estimated values are shown in Table 2.7.

With the hyperparameters from the marginal likelihoods, we can perform Bayesian sensitivity analysis via Bregman divergence both for the frequency priors and the average severity priors. For calculating $D_{\psi}^{R}$, the following convex function $\psi(z) = z \log z - z + 1$ is used, which is a special case of a class of convex functions considered in Eguchi and Kano (2001). As shown in Figure 2.1, sensitivity of $\pi_{\epsilon}(\theta_N|n)/\pi(\theta_N|n)$ is always higher when we used the naive prior than when we used the proposed prior.

Table 2.7: Regression estimates from marginal average severity likelihoods

| | Gamma | | MVGP | |
|---|---|---|---|---|
| | Estimate | Std. Error | Estimate | Std. Error |
| (Intercept) | 8.8954 | 2.2013 | 8.7666 | 1.5147 |
| TypeCity | 2.5348 | 2.0274 | 2.5222 | 1.3807 |
| TypeCounty | 2.5214 | 2.0309 | 2.4038 | 1.3865 |
| TypeSchool | 1.1701 | 2.0645 | 1.2718 | 1.4083 |
| TypeTown | 1.6037 | 2.0552 | 1.5909 | 1.4062 |
| TypeVillage | 1.2629 | 2.0284 | 1.2599 | 1.3793 |
| CoverageIM | 0.0326 | 0.0184 | 0.0346 | 0.0152 |
| lnDeductIM | -0.1653 | 0.1442 | -0.1533 | 0.1056 |
| NoClaimCreditIM | -0.1095 | 0.2775 | -0.0957 | 0.2000 |
| FreqIM | -0.4632 | 0.1004 | -0.4448 | 0.0717 |

In case of the average severity priors, as shown in Figure 2.2, under perturbations with uniform, lognormal, and normal priors, again the naive prior, point mass shows higher sensitivity in all perturbation levels. Therefore, we can claim that in both frequency and severity cases, use of the proposed priors are more robust from possible perturbations of the true prior distribution.

Finally, use of the proposed priors could be justified under out-of-sample validation. Using the predictive distributions in both frequency and the average severity, the expected total losses are calculated based on the observed characteristics of policyholders and compared with the actual total claims in the test set. As shown in Table 2.8, the combination of proposed models in both frequency and the average severity show better performance on the prediction results of total claims in terms of both root-mean-square error (RMSE) and mean absolute error (MAE).

Table 2.8: Validation measures for the prediction of total claims

| | Naive model | Proposed model |
|---|---|---|
| RMSE | 6692.328 | 6443.930 |
| MAE | 1800.494 | 1541.881 |

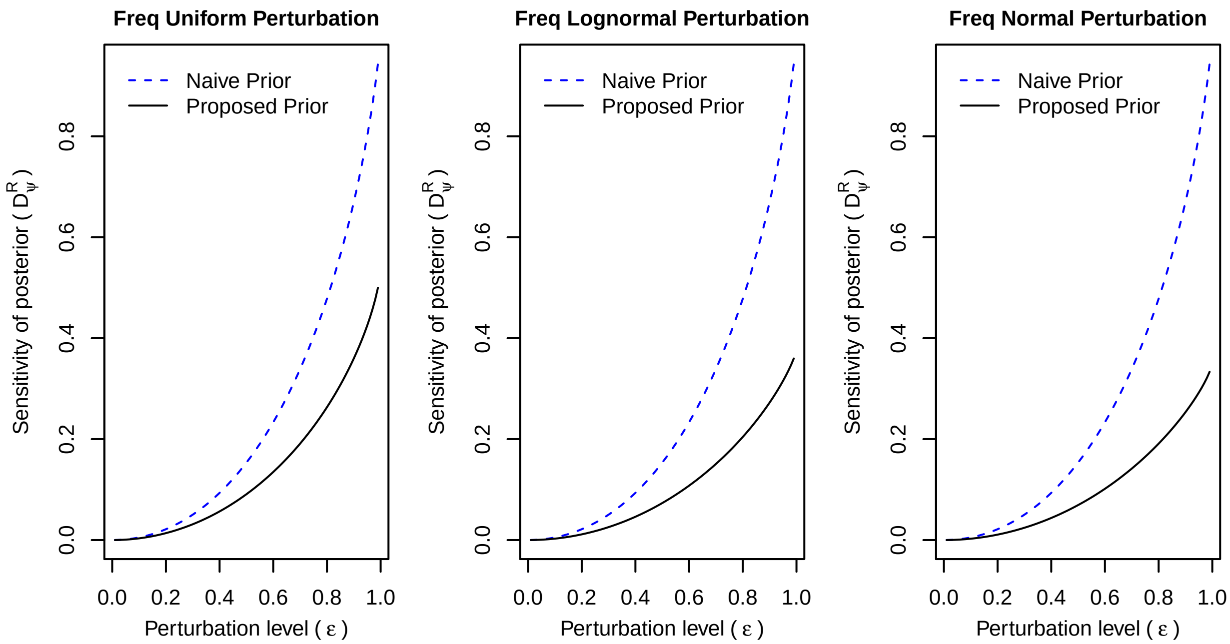



Figure 2.1: Sensitivities of frequency priors with various perturbation priors

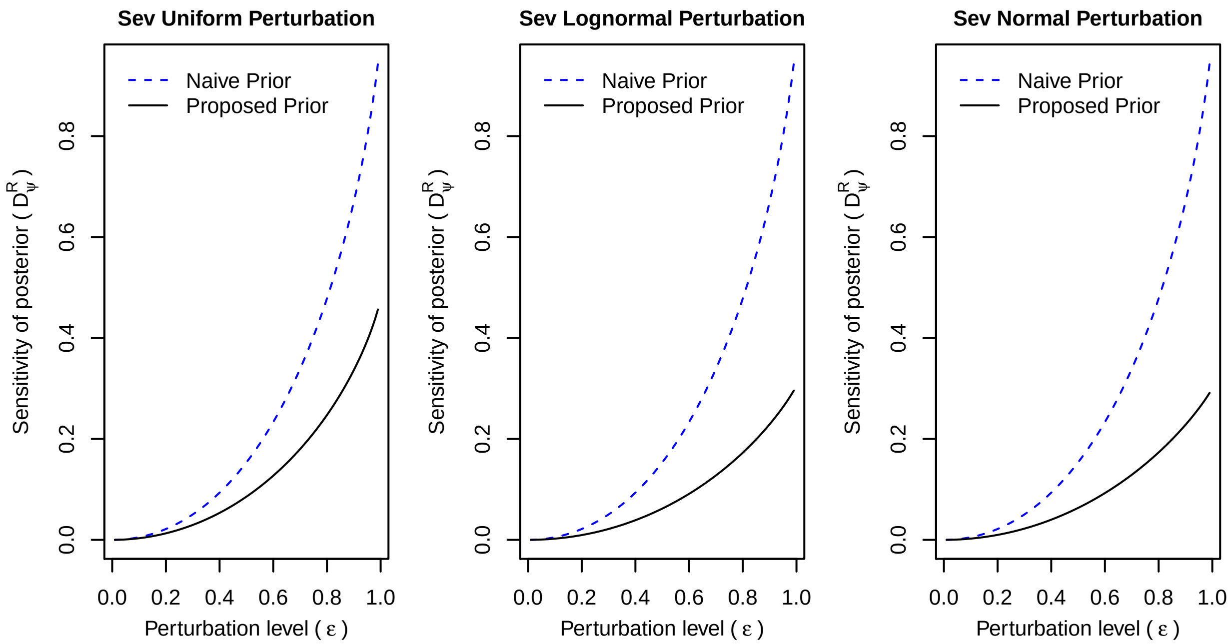



Figure 2.2: Sensitivities of severity priors with various perturbation priors

## 2.4 Conclusion

In this chapter, we explored a justification of using a non-point mass prior in terms of prior elicitation and Bayesian sensitivity analysis. Use of a point mass prior on random effects might be too informative so that we may consider less informative priors on random effects. Upon the use of Bayesian sensitivity analysis with Bregman divergence, it was shown that the proposed priors yield the more robustness than the naive priors both in frequency and the average severity components, respectively. Furthermore, the predicted values of total claims based on the estimates from the proposed marginal likelihood ended up with better performance than the prediceted values of total claims based on the naive independent data likelihood. Therefore, this study provides a theoretical framework to test presence of non-constant random effects in longitudinal insurance claim datasets as well as the empirical results.

# Chapter 3

# Predictive compound risk models with dependence

## 3.1 Introduction

In the actuarial practice of predicting insurance pure premium based on aggregate claims, the two-part model has been widely used. It decomposes aggregate claims into two parts: one for frequency and the other for severity. Due to the apparent non-normality of claim frequency and severity, other types of distribution have been used, and it is noteworthy that most of them are closely related to the use of generalized linear models (GLMs) introduced by Nelder and Wedderburn (1972). With GLMs, we do not need to adhere to ordinary linear regression models, which implicitly assumes normally distributed random errors. Indeed, GLM extends the possible distributions for the regression to members of the exponential family, which includes Gaussian, Poisson, binomial, and gamma. The two-part GLM framework has been commonly used for modeling loss frequency and the loss severity as a benchmark.

Despite its popularity and flexibility, the two-part GLM does not directly address two important concerns. First, the two-part GLM assumes independence between claims frequency and severity usually for ease of modeling and computation. However, we may not preclude the possible dependence between the frequency and severity components. Inspired by this idea, there has been an increasing interest in developing models to capture the possible dependence as in Shi et al. (2015) and Lee et al. (2018). Secondly, it is also important to consider the repeated measurement of claims because it can provide further insights as to the unobserved heterogeneity of the policyholders. For example, consider a situation with two identical groups of policyholders in terms of observed covariates. If one group turns out to have very high claims for 3 consecutive years while the other group has no claim for 3 years, then we may suspect that the observed covariates might not be enough for capturing the risk of each policyholder so that we need to incorporate the concept of random effects on claim occurrence, which is analogous to the unobserved heterogeneity of the riskiness of policyholders. For detailed examples of capturing unobserved heterogeneity via random effects, see Frees and Kim (2006) and Kim et al. (2017).

In this regard, this article explores the benefits of using random effects for predicting insurance claims observed longitudinally, or over a period of time, within a two-part framework, without restricting to the distributions in the exponential family and the assumption of independence between frequency and severity. One can see that Jeong et al. (2020) also considered dependent compound risk model, which specified distributional models within the family of GLMs and used Gaussian random effects. However, our work is distinguished from Jeong et al. (2020) whereby random effects were only used to control for unobserved heterogeneity in risks and obtain better estimates of the fixed effects, or regression coefficients for observed heterogeneity. On the contrary, our work considers the random effects not only in the estimation of fixed effects, but also in the prediction of future claims in the form of crediblity premium.

For the construction of the dependent random effects model in this paper, we propose the use of families of distributions with conjugate random effects, which enables us to obtain explicit moments, marginal likelhood, and predictive distributions, as proposed in Lee and

Nelder (1996) and Molenberghs et al. (2010). For the frequency component, we use Poisson model with gamma random effects. For the severity component, we utilize the family of GB2 distributions derived with an underlying distribution based on generalized gamma (G-gamma) and random effects based on generalized inverse gamma (GI-gamma).

Note that under our proposed dependent compound risk random effects model, we may derive the credibility premium of the compound sum which exploits the longitudinal property as well as adjustments for the possible dependence between the frequency and severity. Credibility premium based on random effects has been explored in the actuarial literature. For example, Frees et al. (1999) provided a general framework which integrates well known credibility models based on the use of linear mixed models. Frangos and Vrontos (2001) suggested the use of conjugate random effects in both the frequency and severity components so that the credibility premium of the compound sum can be expressed as a product of the credibility premiums of the frequency and severity, which implicitly assumes the independence. Shevchenko and Wuthrich (2006) also considered a similar approach to Frangos and Vrontos (2001) in order to model operational risks using Bayesian posterior premium. However, since we cannot preclude the possible dependence between frequency and severity, there has been some research work to incorporate dependence in the calculation of credibility premiums of the compound sum as in Hernández-Bastida et al. (2009) and Gómez-Déniz (2016).

In this article, we propose a dependent compound risk random effects model that enables us to extend both Frangos and Vrontos (2001) and Garrido et al. (2016) and includes them as special cases. We derive the credibility premium of the compound sum as a product of not only the credibility premiums of frequency and severity, but also of an adjustment factor, $D_N(\gamma)$, which accounts for the possible dependence between frequency and severity in a flexible manner. Here, $D_N(\gamma)$ is an informative measure of the strength of the association betweem frequency and severity. For example, if there is a strong and positive association, then $D_N(\gamma)$ is far greater than 1, which implies that the credibility premium for the compound sum should be much larger than the case of independence. On the other hand, if there is weak and negative association, then $D_N(\gamma)$ is slightly less than 1, which implies the opposite. Finally, if there is no dependence between frequency and severity, then $D_N(\gamma)$ is exactly 1, which implies the credibility premium of the compound sum is the product of the credibility premiums of frequency and severity, as in Frangos and Vrontos (2001).

For calibration of the proposed models, we use a longitudinal claims dataset from a Singapore automobile insurance company, which includes policyholder characteristics as well as claim observations for multiple years. This dataset has been used in previous actuarial work, including Antonio and Valdez (2012) and Jeong et al. (2018).

Organization of this paper is as follows. In Section 3.2, we introduce the general concept of the dependent compound risk random effects model and our model specifications for the frequency and average severity components so that our paper is self-contained and provides the notation used throughout the paper. In Section 3.3, we provide our main theoretical results on the derivation of credibility premium of the compound sum without assuming independence. In Section 3.4, we describe our empirical data with some preliminary investigation. In Section 3.5, we analyze the estimation results and introduce model validation methods used in here as well. In the end, in Section 3.6 we conclude this paper with some final remarks. The appendices are included to show the detailed calculations of the credibility premium of the compound sum of claims, as well as both marginal and predictive densities of the average severity for GP and GB2 distributions.

## 3.2 The dependent compound risk random effects model

Garrido et al. (2016) incorporated dependence between loss frequency and severity as follows. According to their framework, we consider an insurer's portfolio as a cross-sectional data (in other words, for a fixed time period which is usually a single year). Here, $N$ refers to the number of claims and the size of claims are denoted by $C_1, C_2, \ldots, C_N$. Then, the total loss can be expressed as

$$S = C_1 + C_2 + \cdots + C_N.$$

Conventionally, when $N = 0$, the total loss $S = 0$ as well. Only in case of $N > 0$, we define the average of loss severity as $\overline{C} = S/N$, which leads to the expression of the aggregate loss as $S = N\overline{C}$. Denote $\mathbf{x} = (x_1, \ldots, x_p)$ as a set of $p$ explanatory variables (covariates); we introduce dependence between the loss frequency and the average of loss severity as follows. If we set link functions $g_N$ and $g_C$ for frequency and severity in GLMs, respectively, then the conditional expectation of the loss frequency and the average of loss severity is given by

$$\nu = \mathbb{E}\left[N|\mathbf{x}\right] = g_N^{-1}(\mathbf{x}\alpha) \qquad \text{and} \qquad \mu_\gamma = \mathbb{E}\left[\overline{C}|\mathbf{x}, N\right] = g_C^{-1}(\mathbf{x}\beta + \gamma N). \tag{3.1}$$

Garrido et al. (2016) provides a good foundation for a two-part dependent GLM, however, their work is limited to cross-sectional data. Indeed, it is more typically for portfolios of general insurance to observe claims in a longitudinal format. In other words, it contains $(N_{it}, C_{itj}, \mathbf{x}_{it}, e_{it})'$ as observations of independent policyholders for calendar year $t = 1, \ldots, T_i$ and for policyholder $i = 1, \ldots, M$. Now we fix a number $T$ so that $T_i \leq T$ and this does not preclude us from allowing unbalanced data. As usual, $\mathbf{x}_{it}$ refers to the covariates which describe characteristics of each policyholder and $e_{it} \in (0, 1]$ refers to the length of exposure of the policyholder within calendar year $t$; in some cases, a policyholder may not have a full exposure for a given calender year.

Furthermore, $N_{it}$ stands for the number of claims and $C_{itj}$ denotes the observed claim size where subscript $j$ is additionally required to distinguish multiple claims that may happen in a calendar year so that $j = 1, \ldots, N_{it}$. For each calendar year $t$, we specify the claim severity distribution by defining $\overline{C}_{it}$ as follows provided $N_{it} > 0$.

$$\overline{C}_{it} = \frac{1}{N_{it}} \sum_{j=1}^{N_{it}} C_{itj}. \tag{3.2}$$

Thus, the joint density of our dependent compound risk random effects model is given as

$$f(n_{it}, \overline{c}_{it} | \theta_i^N, \theta_i^C) = \begin{cases} f_N(n_{it}|\theta_i^N) \times f_{\overline{C}|N}(\overline{c}_{it}|\theta_i^C, n_{it}), & n_{it} > 0 \\ f_N(0|\theta_i^N) & n_{it} = 0 \end{cases} \tag{3.3}$$

where $f_N$ and $f_{\overline{C}|N}$, provided they exist, refer to the density functions for frequency and average severity, respectively. Note that both $\overline{C}$ and $f_{\overline{C}|N}$ are well-defined only in the case of $N > 0$. $\theta_i^N$ and $\theta_i^C$ refer to the random effects for the frequency and average severity of policyholder $i$ with corresponding prior densities $\pi_N$ and $\pi_C$, respectively. The construction in (3.3) is similar to the basic two-part model of frequency and severity, with the exception that our specification allows dependence between frequency and severity as well as the presence of random effects for each policyholder. We also consider dependence between frequency and average severity by using the number of claims $N$ as a linear predictor in the mean function for the average severity as in (3.1).

Now we define the compound sum as

$$S_{it} = \sum_{j=1}^{N_{it}} C_{itj} = N_{it}\overline{C}_{it} \tag{3.4}$$

to refer to the aggregate claims for policyholder $i$ in calendar year $t$. Let us denote $\mathbf{n}_{i,T_i} = (n_{i1}, n_{i2}, \ldots, n_{i,T_i})$ and $\overline{\mathbf{c}}_{i,T_i} = (\overline{c}_{i1}, \overline{c}_{i2}, \ldots, \overline{c}_{i,T_i})$. Note that due to the presence of random effects, predictive distribution of frequency can be expressed as follows:

$$
\begin{aligned}
f(n_{i,T_i+1}|\mathbf{n}_{i,T_i}) &= \frac{f(n_{i1},\ldots,n_{i,T_i},n_{i,T_i+1})}{f(n_{i1},\ldots,n_{i,T_i})} = \frac{\int \left(\prod_{t=1}^{T_i+1} f(n_{it}|\theta)\right)\pi_N(\theta)d\theta}{f(n_{i1},\ldots,n_{i,T_i})} \\
&= \int f(n_{i,T_i+1}|\theta)\frac{\pi_N(\theta)\prod_{t=1}^{T_i} f(n_{it}|\theta)}{f(n_{i1},\ldots,n_{i,T_i})}d\theta = \int f(n_{i,T_i+1}|\theta)\pi_N(\theta|\mathbf{n}_{i,T_i})d\theta.
\end{aligned}
$$

Likewise, predictive distribution of the average severity can be expressed as follows after suppressing conditioning argument on $n$:

$$
\begin{aligned}
f(\overline{c}_{i,T_i+1}|\overline{\mathbf{c}}_{i,T_i}) &= \frac{f(\overline{c}_{i1},\ldots,\overline{c}_{i,T_i},\overline{c}_{i,T_i+1})}{f(\overline{c}_{i1},\ldots,\overline{c}_{i,T_i})} = \frac{\int \left(\prod_{t=1}^{T_i+1} f(\overline{c}_t|\theta)\right)\pi_C(\theta)d\theta}{f(\overline{c}_{i1},\ldots,\overline{c}_{i,T_i})} \\
&= \int f(\overline{c}_{i,T_i+1}|\theta)\frac{\pi_C(\theta)\prod_{t=1}^{T_i+1} f(\overline{c}_t|\theta)}{f(\overline{c}_{i1},\ldots,\overline{c}_{i,T_i})}d\theta = \int f(\overline{c}_{i,T_i+1}|\theta)\pi_C(\theta|\overline{\mathbf{c}}_{i,T_i})d\theta.
\end{aligned}
$$

Based on these predictive distributions, the predictive mean of $S_{i,T_i+1}$ can be expressed as follows:

$$
\mathbb{E}\left[S_{i,T_i+1}|\mathbf{n}_{i,T_i},\overline{\mathbf{c}}_{i,T_i}\right] = \mathbb{E}\left[N_{i,T_i+1}\overline{C}_{i,T_i+1}|\mathbf{n}_{i,T_i},\overline{\mathbf{c}}_{i,T_i}\right] = \mathbb{E}\left[N_{i,T_i+1}\mathbb{E}\left[\overline{C}_{i,T_i+1}|N_{i,T_i+1}\right]|\mathbf{n}_{i,T_i},\overline{\mathbf{c}}_{i,T_i}\right].
$$

If we assume that $\mathbb{E}\left[\overline{C}_{it}|N_{it}\right] = \hat{\mu}_{it}\, e^{\gamma N_{it}}$ such that $\hat{\mu}_{it}$ is independent of $N_{it}$, then we obtain the following predictive mean of $S_{i,T_i+1}$:

$$
\begin{aligned}
\mathbb{E}\left[S_{i,T_i+1}|\mathbf{n}_{i,T_i},\overline{\mathbf{c}}_{i,T_i}\right] &= \mathbb{E}\left[N_{i,T_i+1}\overline{C}_{i,T_i+1}|\mathbf{n}_{i,T_i},\overline{\mathbf{c}}_{i,T_i}\right] = \mathbb{E}\left[N_{i,T_i+1}\mathbb{E}\left[\overline{C}_{i,T_i+1}|N_{i,T_i+1}\right]|\mathbf{n}_{i,T_i},\overline{\mathbf{c}}_{i,T_i}\right] \\
&= \hat{\mu}_{i,T_i+1}\times\mathbb{E}\left[N_{i,T_i+1}e^{\gamma N_{i,T_i+1}}|\mathbf{n}_{i,T_i}\right].
\end{aligned} \tag{3.5}
$$

### 3.2.1 Frequency model specification

For frequency $N_{it}$, we have the following two candidate models where we denote $\nu_{it} = e^{\mathbf{x}_{it}\alpha}$ with $\alpha$ as a $p\times 1$ parameter vector for the covariates associated with frequency for which they may be different from that of the average severity. Here we account for the exposure $e_{it}$ and this can be done by incorporating $e_{it}$ as an offset to the mean parameter $\nu_{it}$.

(1) Poisson GLM: $N_{it} \sim \mathcal{P}(\nu_{it})$

(2) Poisson/gamma random effects model (= multivariate negative binomial model)

Now let us explain the Poisson/gamma random effects model, which is given as

$$N_{it}|\theta_i^N \sim \ \mathcal{P}(\nu_{it}\theta_i^N) \text{ and } \theta_i^N \sim \ \mathcal{G}(r, 1/r) \tag{3.6}$$

so that $\mathbb{E}\left[N_{it}|\theta_i^N\right] = \text{Var}\left[N_{it}|\theta_i^N\right] = \nu_{it}\theta_i^N$. From above specification, we see that $N_{it} \sim \mathcal{NB}\left(r, \dfrac{\nu_{it}}{r+\nu_{it}}\right)$ and the following relations hold:

$$\begin{aligned}
f_N(n) &= \binom{n+r-1}{n} \left(\frac{r}{r+\nu_{it}}\right)^r \left(\frac{\nu_{it}}{r+\nu_{it}}\right)^n, \\
\mathbb{E}\left[N_{it}\right] &= \mathbb{E}\left[\mathbb{E}\left[N_{it}|\theta_i^N\right]\right] = \mathbb{E}\left[\theta_i^N \nu_{it}\right] = \nu_{it}, \\
\text{Var}\left[N_{it}\right] &= \text{Var}\left[\mathbb{E}\left[N_{it}|\theta_i^N\right]\right] + \mathbb{E}\left[\text{Var}\left[N_{it}|\theta_i^N\right]\right] = \nu_{it}\left(1+\frac{\nu_{it}}{r}\right).
\end{aligned} \tag{3.7}$$

Although here we have $N_{it} \sim \mathcal{NB}\left(r, \dfrac{\nu_{it}}{r+\nu_{it}}\right)$, our model specification is different from the usual negative binomial GLM for frequency component because $N_{it}$ is not marginally independent. Let $\pi_N(\theta)$ be the probability density with respect to the gamma distribution. Then according to Boucher et al. (2008), it is known that the joint density for the multi-year claim counts is given as follows, which is called the multivariate negative binomial (MVNB) distribution:

$$\begin{aligned}
f_{N_i}(n_{i1}, \ldots, n_{iT_i}) &= \int \prod_{t=1}^{T_i} f_{N|\theta^N}(n_{it}|\theta)\pi_N(\theta)d\theta \\
&= \prod_{t=1}^{T_i} \left(\frac{e^{\mathbf{x}_{it}\alpha}}{\sum_{t=1}^{T_i} e^{\mathbf{x}_{it}\alpha} + r}\right)^{n_{it}} \left(\frac{r}{\sum_{t=1}^{T_i} e^{\mathbf{x}_{it}\alpha} + r}\right)^r \frac{\Gamma\left(\sum\limits_{t=1}^{T_i} n_{it} + r\right)}{\Gamma(r)\prod\limits_{t=1}^{T_i} n_{it}!}.
\end{aligned}$$

Note that MVNB distributions have been widely used in the actuarial literature including, but not limited to, first proposed by Hausman et al. (1984) and subsequently followed by

Dionne and Vanasse (1989) and Shi and Valdez (2014). According to Winkelmann (2008), the MVNB distribution described in (3.6) is referred as MVNB-II type. A broad characterization of MVNB distributions can be found in Doss (1979).

With derived joint density of MVNB distribution, we estimate $\hat{\alpha}$ and $\hat{r}$ by maximizing the following likelihood function:

$$\begin{aligned}
\ell_N &= \sum_{i=1}^{M} \left( \log \int \prod_{t=1}^{T_i} f_{N|\theta^N}(n_{it}|\theta)\pi_N(\theta)d\theta \right) \\
&= \sum_{i=1}^{M}\sum_{t=1}^{T_i} \left[n_{it}\mathbf{x}_{it}\alpha - \log\Gamma(n_{it}+1)\right] + \sum_{i=1}^{M} \log\Gamma\left(\sum_{t=1}^{T_i} n_{it} + r\right) - \\
&\quad \sum_{i=1}^{M}\left[\left(\sum_{t=1}^{T_i} n_{it} + r\right)\log\left(\sum_{t=1}^{T_i} e^{\mathbf{x}_{it}\alpha} + r\right)\right] + M\left[r\log r - \log\Gamma(r)\right].
\end{aligned} \tag{3.8}$$

Finally, it is not difficult to show that the predictive distribution of $N_{i,T_i+1}$ given $\mathbf{n}_{i,T_i}$ still has a negative binomial distribution as follows:

**Lemma 3.1.** *Suppose $N_{i1}, N_{i2}, \ldots, N_{it}$ follows MVNB distribution as defined in (3.6). Then*

$$N_{i,T_i+1}|\mathbf{n}_T \sim \mathcal{NB}\left(r_{i,T_i}, \frac{\nu_{i,T_i+1}}{\tilde{r}_{i,T_i} + \nu_{i,T_i+1}}\right),$$

*where $r_{i,T_i} = r + \sum_{t=1}^{T_i} n_{it}$ and $\tilde{r}_{i,T_i} = r + \sum_{t=1}^{T_i} \nu_{it}$.*

*Proof.* From (3.7), we see that if $N_{it}|\theta_i^N \sim \mathcal{P}(\nu_{it}\theta_i^N)$ and $\theta_i^N \sim \mathcal{G}(r, 1/r)$, then $N_{it} \sim \mathcal{NB}\left(r, \dfrac{\nu_{it}}{r+\nu_{it}}\right)$. Furthermore, it is easy to show that $\theta_i^N|\mathbf{n}_{i,T_i} \sim \mathcal{G}(r_{i,T_i}, 1/\tilde{r}_{i,T_i})$ because

$$\begin{aligned}
\pi_N(\theta|\mathbf{n}_{i,T_i}) &\propto \pi_N(\theta)\prod_{t=1}^{T_i} f(n_{it}|\theta) \propto \theta^{r-1}\exp(-r\theta)\prod_{t=1}^{T_i}\theta^{n_{it}}\exp(-\nu_{it}\theta) \\
&\propto \theta^{\sum_{t=1}^{T_i} n_{it}+r-1}\exp\left(-\left[r + \sum_{t=1}^{T_i}\nu_{it}\right]\theta\right) = \theta^{r_{i,T_i}-1}\exp\left(-\tilde{r}_{i,T_i}\theta\right).
\end{aligned}$$

Therefore, $N_{i,T_i+1}|\mathbf{n}_{i,T_i} \sim \mathcal{NB}\left(r_{i,T_i}, \dfrac{\nu_{i,T_i+1}}{\tilde{r}_{i,T_i}+\nu_{i,T_i+1}}\right)$ with $\mathbb{E}\left[N_{i,T_i+1}|\mathbf{n}_{i,T_i}\right] = \dfrac{r_{i,T_i}}{\tilde{r}_{i,T_i}}\nu_{i,T_i+1} =$

$\dfrac{r+\sum_{t=1}^{T_i} n_{it}}{r+\sum_{t=1}^{T_i} \nu_{it}} \nu_{i,T_i+1}.$ $\square$

### 3.2.2 Severity model specification

For average severity $\overline{C}_{it}|N_{it}$, we have four candidate models where we denote $\mu_{it} = e^{\mathbf{x}_{it}\beta + N_{it}\gamma}$ with $\beta$ as a $p \times 1$ parameter vector for the covariates associated with average severity and $\gamma$ is parameter which is used to measure the dependency between frequency and average severity.

(1) GB2 model: $\overline{C}_{it}|N_{it} \overset{indep}{\sim} \mathcal{GB}2\left(k+1,\, \mu_{it}\dfrac{\Gamma(k+1)}{\Gamma(k+1-1/p)}\dfrac{\Gamma(N_{it}/\phi)}{\Gamma(N_{it}/\phi+1/p)},\, \dfrac{N_{it}}{\phi},\, p\right)$ so that $\mathbb{E}\left[\overline{C}_{it}|N_{it}\right] = \mu_{it}$.

(2) Gamma/Normal random effects model (= Gamma GLMM): $\overline{C}_{it}|N_{it}, \theta_i^C \sim \mathcal{G}\left(\dfrac{N_{it}}{\phi}, \theta_i^C \mu_{it}\dfrac{\phi}{N_{it}}\right)$ so that $\mathbb{E}\left[\overline{C}_{it}|N_{it}, \theta_i^C\right] = \theta_i^C \mu_{it}$ and $\dfrac{\mathrm{Var}\left[\overline{C}_{it}|N_{it}, \theta_i^C\right]}{\mathbb{E}\left[\overline{C}_{it}|N_{it}, \theta_i^C\right]^2} = \dfrac{\phi}{N_{it}}$, where $\log \theta_i^C \sim \mathcal{N}\left(-\dfrac{\sigma^2}{2},\, \sigma^2\right)$ so that $\mathbb{E}\left[\theta_i^C\right] = 1$.

(3) Gamma/Inv-gamma random effects model (= MVGP model)

(4) G-gamma/GI-gamma random effects model (= MVGB2 model)

It is well-known that most of general insurance claim datasets show heavy right-tail behavior. In this regard, we consider GB2 model as a benchmark, which might fit better than light-tail distributions such as gamma. Since our proposed severity models are MVGP model and MVGB2 model, let us examine them more carefully. First, gamma/inv-gamma random effects model is given as follows:

$$\overline{C}_{it}|N_{it}, \theta_i^C \sim \ \mathcal{G}\left(\frac{N_{it}}{\phi}, \theta_i^C \mu_{it}\frac{\phi}{N_{it}}\right) \text{ and } \theta_i^C \sim \mathcal{IG}(k+1, k)$$

so that $\mathbb{E}\left[\bar{C}_{it}|N_{it}, \theta_i^C\right] = \theta_i^C \mu_{it}, \mathrm{Var}\left[\bar{C}_{it}|N_{it}, \theta_i^C\right] = (\theta_i^C)^2 \mu_{it}^2 \dfrac{\phi}{N_{it}}$. From the above specification, we see that $\bar{C}_{it}|N_{it}$ follows a generalized Pareto (GP) distribution, in other words, $\bar{C}_{it}|N_{it} \sim$

$\mathcal{GP}\left(k+1, \mu_{it}k\dfrac{\phi}{N_{it}}, \dfrac{N_{it}}{\phi}\right)$ and the following relations hold:

$$
\begin{aligned}
\mathbb{E}\left[\bar{C}_{it}|N_{it}\right] &= \mathbb{E}\left[\mathbb{E}\left[\bar{C}_{it}|N_{it}, \theta_i^C\right]\right] = \mathbb{E}\left[\theta_i^C \mu_{it}\right] = \mu_{it}, \\
\mathrm{Var}\left[\bar{C}_{it}|N_{it}\right] &= \mathrm{Var}\left[\mathbb{E}\left[\bar{C}_{it}|N_{it}, \theta_i^C\right]\right] + \mathbb{E}\left[\mathrm{Var}\left[\bar{C}_{it}|N_{it}, \theta_i^C\right]\right] = \frac{\mu_{it}^2}{k-1}\left(1+\frac{k\phi}{N_{it}}\right).
\end{aligned}
$$

Note that generalized Pareto distribution is widely used in ratemaking application by itself, as mentioned in Klugman et al. (2012).

Let $\pi_C(\theta)$ be the probability density with respect to inverse gamma. Then we get the marginal likelihood for the average severity as follows, which can be called multivariate generalized Pareto (MVGP) distribution:

$$
\begin{aligned}
f_{\overline{C_i}|N_i}(\overline{c}_{i1}, \ldots, \overline{c}_{i,T_i}|n_i) &= \int \prod_{t=1}^{T_i} f_{\overline{C}|N,\theta^C}(\overline{c}_{it}|n_{it}, \theta)\pi_C(\theta)d\theta \\
&= \frac{k^{k+1}\prod_{t=1}^{T_i}\left(n_{it}\overline{c}_{it}e^{-\mathbf{x}_{it}\beta - n_{it}\gamma}/\phi\right)^{n_{it}/\phi}}{\left(k+\sum_{t=1}^{T_i} n_{it}\overline{c}_{it}e^{-\mathbf{x}_{it}\beta-n_{it}\gamma}/\phi\right)^{\sum_{t=1}^{T_i} n_{it}/\phi+k+1}} \times \frac{\Gamma\left(\sum_{t=1}^{T_i} n_{it}/\phi+k+1\right)\prod_{t=1}^{T_i}\overline{c}_{it}^{-1}}{\Gamma(k+1)\prod_{t=1}^{T_i}\Gamma(n_{it}/\phi)}.
\end{aligned}
$$

The details of the derivation is provided in Appendix B. Now using the joint density function for $\overline{c}_{i1}, \ldots, \overline{c}_{i,T_i}|\mathbf{n}_{T_i}$, we estimate $\hat{\phi}, \hat{\beta}, \hat{\gamma}$ and $\hat{k}$ by maximizing the following likelihood function:

$$
\begin{aligned}
\ell_{\overline{C}|N} &= \sum_{i=1}^{M}\left(\log \int \prod_{t=1}^{T_i} f_{\overline{C}|N,\theta^C}(\overline{c}_{it}|n_{it},\theta)\pi_C(\theta)d\theta\right) \\
&= \sum_{i=1}^{M}\sum_{t=1}^{T_i}\left[-\log\Gamma(n_{it}/\phi) - \log\overline{c}_{it}\right] + \sum_{i=1}^{M}\log\Gamma\left(\sum_{t=1}^{T_i} n_{it}/\phi+k+1\right) \\
&\quad + \sum_{i=1}^{M}\sum_{t=1}^{T_i} n_{it}/\phi(\log n_{it}\overline{c}_{it} - \mathbf{x}_{it}\beta - n_{it}\gamma - \log\phi) \\
&\quad - \sum_{i=1}^{M}\left[\left(\sum_{t=1}^{T_i} n_{it}/\phi+k+1\right)\log\left(k+\sum_{t=1}^{T_i} n_{it}\overline{c}_{it}e^{-\mathbf{x}_{it}\beta-n_{it}\gamma}/\phi\right)\right] + M\left[(k+1)\log k - \log\Gamma(k+1)\right].
\end{aligned}
\tag{3.9}
$$

Note that generalized Pareto distribution described in (B.1), in Appendix B, is distinguished from that in Rootzén et al. (2006) although they share the same name. In their formulation, the density function of a generalized Pareto distribution is given as

$$f(y) = \frac{\alpha}{\xi}\left(1+\frac{y-\mu}{\xi}\right)^{\alpha-1} = \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)\Gamma(1)}\frac{\xi^{\alpha}}{(y-\mu+\xi)^{\alpha+1}},$$

which is equivalent to $Y-\mu \sim \mathcal{GP}(\alpha, \xi, \tau=1)$ in our formulation described in (B.1). Major differences between the MVGP distributions of ours and Rootzén et al. (2006) are the marginalization and conditioning properties. In our specification, any lower-dimensional marginal or conditional distributions of MVGP distributions remain in the same family, while such property does not exist with MVGP distribution proposed in Rootzén et al. (2006). For details on the marginalization and conditioning properties of our proposed MVGP distribution, see Sections 4 and 6 of Jeong and Valdez (2020a).

Likewise, by denoting $w = \dfrac{\Gamma(k+1)}{\Gamma(k+1-1/p)}$, G-gamma/GI-gamma random effects model is given as follows:

$$\overline{C}_{it}|N_{it}, \theta_i^C \sim \ \mathcal{GG}\left(\frac{N_{it}}{\phi}, \theta_i^C \mu_{it}\frac{\Gamma(N_{it}/\phi)}{\Gamma(N_{it}/\phi+1/p)}, p\right) \ \text{and} \ \ \theta_i^C \sim \mathcal{GIG}(k+1, w, p)$$

so that $\mathbb{E}\left[\bar{C}_{it}|N_{it}, \theta_i^C\right] = \theta_i^C\mu_{it}$ and $\text{Var}\left[\bar{C}_{it}|N_{it}, \theta_i^C\right] = (\theta_i^C)^2\mu_{it}^2\left(\dfrac{\Gamma(N_{it}/\phi+2/p)\Gamma(N_{it}/\phi)}{\Gamma(N_{it}/\phi+1/p)^2}-1\right)$. From the above specification, we see that $\bar{C}_{it}|N_{it}$ follows a generalized beta of the second-kind (GB2) distribution, in other words, $\bar{C}_{it}|N_{it} \sim \mathcal{GB}2\left(k+1,\ w\mu_{it}\dfrac{\Gamma(N_{it}/\phi)}{\Gamma(N_{it}/\phi+1/p)},\ \dfrac{N_{it}}{\phi},\ p\right)$ and the following relations hold:

$$
\begin{aligned}
\mathbb{E}\left[\bar{C}_{it}|N_{it}\right] &= \mathbb{E}\left[\mathbb{E}\left[\bar{C}_{it}|N_{it},\theta_i^C\right]\right] = \mathbb{E}\left[\theta_i^C\mu_{it}\right] = \mu_{it}, \\
\mathrm{Var}\left[\bar{C}_{it}|N_{it}\right] &= \mathrm{Var}\left[\mathbb{E}\left[\bar{C}_{it}|N_{it},\theta_i^C\right]\right] + \mathbb{E}\left[\mathrm{Var}\left[\bar{C}_{it}|N_{it},\theta_i^C\right]\right] \\
&= \mathrm{Var}\left[\theta_i^C\mu_{it}\right] + \mathbb{E}\left[(\theta_i^C)^2\mu_{it}^2\left(\frac{\Gamma(N_{it}/\phi+2/p)\Gamma(N_{it}/\phi)}{\Gamma(N_{it}/\phi+1/p)^2}-1\right)\right] \\
&= \mu_{it}^2\left[\frac{\Gamma(k+1-\frac{2}{p})\Gamma(k+1)}{\Gamma(k+1-\frac{1}{p})^2}\frac{\Gamma(\frac{N_{it}}{\phi}+\frac{2}{p})\Gamma(\frac{N_{it}}{\phi})}{\Gamma(\frac{N_{it}}{\phi}+\frac{1}{p})^2}-1\right].
\end{aligned}
$$

Note that in both cases of GB2 and MVGB2 models, we have the same marginal distribution of $\overline{C}_{it}|N_{it}$. However, while GB2 model assumes marginal independence of $\bar{C}_{it}$ without considering possible longitudinality, MVGB2 model assumes $\bar{C}_{it}$ are not marginally independent, but only independent conditional on individual unobserved heterogeneity $\theta_i^C$.

Let $\pi_C(\theta)$ be the probability density with respect to generalized inverse gamma. Then we get the following marginal likelihood for the average severity, which is called multivariate generalized beta-II (MVGB2) distribution proposed by Yang et al. (2011):

$$
\begin{aligned}
f_{\overline{C_i}|N_i}(\overline{c}_{i1},\ldots,\overline{c}_{i,T_i}|n_i) &= \int\prod_{t=1}^{T_i} f_{\overline{C}|N,\theta^C}(\overline{c}_{it}|n_{it},\theta)\pi_C(\theta)d\theta \\
&= \frac{w^{pk+p}\prod_{t=1}^{T_i}\left(\overline{c}_{it}z_{it}e^{-\mathbf{x}_{it}\beta-n_{it}\gamma}\right)^{pn_{it}/\phi}}{\left(w^p+\sum_{t=1}^{T_i}(\overline{c}_{it}z_{it}e^{-\mathbf{x}_{it}\beta-n_{it}\gamma})^p\right)^{\sum_{t=1}^{T_i}n_{it}/\phi+k+1}} \times \frac{\Gamma\left(\sum_{t=1}^{T_i}n_{it}/\phi+k+1\right)\prod_{t=1}^{T_i}\overline{c}_{it}^{-1}p^{T_i}}{\Gamma(k+1)\prod_{t=1}^{T_i}\Gamma(n_{it}/\phi)}
\end{aligned}
$$

Considered a flexible parametric distribution with four parameters describing scale and various shapes, this gives a natural association structure within claims of a policyholder $i$ as well. The details of the derivation is provided in Appendix B. One can see that apart from flexibility, the family of GB2 distributions has various advantages, which includes explicit moments and joint likelihood, better fit on the tail, and relationships with other well known distributions. For example, when the power parameter $p$ equals to 1 in GB2, then it is reduced to the aforementioned generalized Pareto (GP) distribution.

Now using the joint density function for $\overline{c}_{i1}, \dots, \overline{c}_{i,T_i}|n_i$, we estimate $\hat{\phi}, \hat{\beta}, \hat{\gamma}, \hat{p}$ and $\hat{k}$ by maximizing the following likelihood function:

$$
\begin{aligned}
\ell_{\overline{C}|N} &= \sum_{i=1}^{M} \left( \log \int \prod_{t=1}^{T_i} f_{\overline{C}|N,\theta^C}(\overline{c}_{it}|n_{it}, \theta)\pi_C(\theta) d\theta \right) \\
&= \sum_{i=1}^{M} \left[ \sum_{t=1}^{T_i} \left( -\log \Gamma\left(\frac{n_{it}}{\phi}\right) - \log\left(\frac{\overline{c}_{it}}{p}\right)\right) + \log \Gamma\left(\sum_{t=1}^{T_i} \frac{n_{it}}{\phi} + k + 1\right)\right] \\
&\quad + p \sum_{i=1}^{M} \sum_{t=1}^{T_i} n_{it}/\phi (\log \overline{c}_{it} z_{it} - \mathbf{x}_{it}\beta - n_{it}\gamma) \\
&\quad - \sum_{i=1}^{M} \left[ \left( \sum_{t=1}^{T_i} n_{it}/\phi + k + 1 \right) \log \left( w^p + \sum_{t=1}^{T_i} (\overline{c}_{it} z_{it} e^{-\mathbf{x}_{it}\beta - n_{it}\gamma})^p \right) \right] \\
&\quad + M \left[ (k+1) p \log w - \log \Gamma(k+1) \right],
\end{aligned}
\tag{3.10}
$$

where $w = \dfrac{\Gamma(k+1)}{\Gamma(k+1-1/p)}$ and $z_{it} = \dfrac{\Gamma(n_{it}/\phi + 1/p)}{\Gamma(n_{it}/\phi)}$.

## 3.3 Credibility premium of compound sum with dependence

In this section, we derive credibility premium for compound sum when we relax the assumption of independence between frequency and severity. As already alluded earlier, it is important to consider the possible dependence in order to further account for the unobserved heterogeneity among the policyholders. We show that our proposed dependent two-part random effects model enables us to obtain credibility premium of $S_{i,T_i+1}$, given information on $\mathbf{n}_{i,T_i}$, $\overline{\mathbf{c}}_{i,T_i}$. The results are summarized in the following theorem. For notational simplicity, we suppress subscript $i$ throughout this section and thereafter, although the proposed credibility premium is indeed estimated for each policyholder.

**Theorem 3.1.** *Suppose $(N_1, N_2, \dots, N_t)$ follows MVNB distribution as defined in (3.6). Moreover, let us assume that $\mathbb{E}\left[\overline{C}_t|n_t\right] = \theta^C e^{\mathbf{x}_t\beta} e^{n_t\gamma} = \theta^C \tilde{\mu}_t e^{n_t\gamma}$. Then, the credibility* *premium of $S_{T+1}$ is given as follows:*

$$\begin{aligned}
\mathbb{E}\left[S_{T+1}|\mathbf{n}_T,\overline{\mathbf{c}}_T\right] &= \mathbb{E}\left[\theta^C|\mathbf{n}_T,\overline{\mathbf{c}}_T\right]\, \tilde{\mu}_{T+1} \times \frac{r+\sum_{t=1}^T n_t}{r+\sum_{t=1}^T \nu_t}\nu_{T+1} \times e^{\gamma}\left[1-\left(\frac{\nu_{T+1}}{\tilde{r}_T}\right)(e^{\gamma}-1)\right]^{-r_T-1} \\
&= \mathbb{E}\left[\tilde{\mu}_{T+1}\theta^C|\mathbf{n}_T,\overline{\mathbf{c}}_T\right] \times \mathbb{E}\left[\nu_{T+1}\theta^N|\mathbf{n}_T\right] \times D_N(\gamma).
\end{aligned} \tag{3.11}$$

*Proof.* The proof is provided in Appendix A. □

The following corollary is the result suggested by Frangos and Vrontos (2001) for the case of independence.

**Corollary 3.1.** *If $\gamma = 0$, then (3.11) reduces to:*

$$\mathbb{E}\left[S_{T+1}|\mathbf{n}_T,\overline{\mathbf{c}}_T\right] = \mathbb{E}\left[\theta^C|\mathbf{n}_T,\overline{\mathbf{c}}_T\right]\, \tilde{\mu}_{T+1} \times \frac{r+\sum_{t=1}^T n_t}{r+\sum_{t=1}^T \nu_t}\nu_{T+1} = \mathbb{E}\left[\mu_{T+1}\theta^C|\mathbf{n}_T,\overline{\mathbf{c}}_T\right] \times \mathbb{E}\left[\nu_{T+1}\theta^N|\mathbf{n}_T\right],$$

Note that if $r \to \infty$, then MVNB distribution defined in (3.6) converges to the naive independent Poisson model and the formula for the credibility premium of $S_{T+1}$ in (3.11) reduces to:

$$\begin{aligned}
\mathbb{E}\left[S_{T+1}|\mathbf{n}_T,\overline{\mathbf{c}}_T\right] &= \mathbb{E}\left[\theta^C|\mathbf{n}_T,\overline{\mathbf{c}}_T\right]\, \tilde{\mu}_{T+1} \times \lim_{r\to\infty}\left\{\frac{r+\sum_{t=1}^T n_t}{r+\sum_{t=1}^T \nu_t}\nu_{T+1} \times e^{\gamma}\left[1-\left(\frac{\nu_{T+1}}{\tilde{r}_T}\right)(e^{\gamma}-1)\right]^{-r_T-1}\right\} \\
&= \mathbb{E}\left[\tilde{\mu}_{T+1}\theta^C|\mathbf{n}_T,\overline{\mathbf{c}}_T\right] \times \nu_{T+1} \times \exp\left(\nu_{T+1}(e^{\gamma}-1)+\gamma\right).
\end{aligned}$$

According to Theorem 3.1, the credibility premium of $S_{T+1}$ can be expressed as the product of credibility premium of $N_{T+1}$, credibility premium of $\overline{C}_{T+1}$, and adjustment factor $D_N(\gamma)$, which helps us explain the dependence between frequency and the average severity. Therefore, we may obtain expressions of $\mathbb{E}\left[S_{T+1}|\mathbf{n}_T,\overline{\mathbf{c}}_T\right]$ in all the four models for the average severity component, described in section 3.2.2, as follows:

**Corollary 3.2.** *Suppose $(N_1, N_2, \ldots, N_t)$ follows MVNB distribution as defined in (3.6). If the average severity component follows GB2 model, then the credibility premium of $S_{T+1}$ is* *given as follows:*

$$\mathbb{E}\left[S_{T+1}|\mathbf{n}_T, \overline{\mathbf{c}}_T\right] = \tilde{\mu}_{T+1} \times \frac{r + \sum_{t=1}^{T} n_t}{r + \sum_{t=1}^{T} \nu_t} \nu_{T+1} \ \times e^{\gamma} \left[1 - \left(\frac{\nu_{T+1}}{\tilde{r}_T}\right)(e^{\gamma} - 1)\right]^{-r_T - 1}.$$

*If the average severity component follows Gamma GLMM, then we have*

$$\mathbb{E}\left[S_{T+1}|\mathbf{n}_T, \overline{\mathbf{c}}_T\right] \approx \frac{\sum_{j=1}^{J} \hat{\theta}_j \, f(\overline{\mathbf{c}}_T|\mathbf{n}_T, \hat{\theta}_j)}{\sum_{j=1}^{J} f(\overline{\mathbf{c}}_T|\mathbf{n}_T, \hat{\theta}_j)} \tilde{\mu}_{T+1} \times \frac{r + \sum_{t=1}^{T} n_t}{r + \sum_{t=1}^{T} \nu_t} \nu_{T+1} \times e^{\gamma} \left[1 - \left(\frac{\nu_{T+1}}{\tilde{r}_T}\right)(e^{\gamma} - 1)\right]^{-r_T - 1},$$

*where $\hat{\theta}_j$'s are generated from $\mathcal{LN}\left(-\sigma^2/2, \, \sigma^2\right)$.*

*If the average severity component follows MVGP model, then we have*

$$\mathbb{E}\left[S_{T+1}|\mathbf{n}_T, \overline{\mathbf{c}}_T\right] = \frac{k\phi + \sum_{t=1}^{T} S_t/\mu_t}{k\phi + \sum_{t=1}^{T} n_t} \tilde{\mu}_{T+1} \times \frac{r + \sum_{t=1}^{T} n_t}{r + \sum_{t=1}^{T} \nu_t} \nu_{T+1} \ \times e^{\gamma} \left[1 - \left(\frac{\nu_{T+1}}{\tilde{r}_T}\right)(e^{\gamma} - 1)\right]^{-r_T - 1}.$$

*Finally, if the average severity component follows MVGB2 model, then we have*

$$\begin{aligned}\mathbb{E}\left[S_{T+1}|\mathbf{n}_T, \overline{\mathbf{c}}_T\right] = &\left(\sum_{t=1}^{T} \left(\frac{S_t}{\mu_t} \frac{\Gamma(n_t/\phi + 1/p)}{\phi\Gamma(n_t/\phi + 1)}\right)^p + w^p\right)^{1/p} \frac{\Gamma(k_T + 1 - 1/p)}{\Gamma(k_T + 1)} \tilde{\mu}_{T+1} \times \\ &\frac{r + \sum_{t=1}^{T} n_t}{r + \sum_{t=1}^{T} \nu_t} \nu_{T+1} \ \times e^{\gamma} \left[1 - \left(\frac{\nu_{T+1}}{\tilde{r}_T}\right)(e^{\gamma} - 1)\right]^{-r_T - 1}.\end{aligned}$$

*Proof.* The proof is provided in Appendix A. □

It should be noted that in the case of Gamma GLMM, the random effects is not a conjugate prior so that the credibility premium is not in explicit form but is expressed as an approximation. Furthermore, note that the credibility premium formula for $S_{T+1}$ in (3.11) is very flexible that provides a natural extension of those derived in previous literature such as in Frangos and Vrontos (2001) and Garrido et al. (2016). For example, for a new policyholder where $T = 0$, there is clearly no available claim history. In this case, (3.11) is reduced to

$$\mathbb{E}\left[S_{0+1}|\mathbf{n}_0, \overline{\mathbf{c}}_0\right] = \mathbb{E}\left[S_1\right] = \mathbb{E}\left[\theta^C \tilde{\mu}_1\right] \times \mathbb{E}\left[N_{T+1} e^{N_1 \gamma}\right] = \tilde{\mu}_1 \ \times \nu_1 \ \times e^{\gamma} \left[1 - \left(\frac{\nu_1}{r}\right)(e^{\gamma} - 1)\right]^{-r-1},$$

which is further reduced to $\tilde{\mu}_1\nu_1 \exp\left(\nu_1(e^\gamma - 1) + \gamma\right)$ as $r \to \infty$. On the other hand, for a policyholder observed for full $T$ years but with no claims, then $\mathbb{E}\left[S_{T+1}|\mathbf{n}_T, \overline{\mathbf{c}}_T\right] = \mathbb{E}\left[S_{T+1}|\mathbf{n}_T = 0\right]$ so that

$$\mathbb{E}\left[S_{T+1}|\mathbf{n}_T, \overline{\mathbf{c}}_T\right] = \mathbb{E}\left[S_{T+1}|\mathbf{n}_T = 0\right] = \tilde{\mu}_{T+1}\frac{r}{r + \sum_{t=1}^T \nu_t}\nu_{T+1} \times e^\gamma \left[1 - \left(\frac{\nu_{T+1}}{\tilde{r}_T}\right)(e^\gamma - 1)\right]^{-r_T - 1}.$$

This can be interpreted that the policyholder is guaranteed premium discount precisely because of a clearly favorable claim history. Finally, in our formulation of credibility premium, $\gamma$ is treated and estimated as a parameter so that the same value of $\gamma$ is used for every policyholder, similar to the regression coefficients $\alpha$ and $\beta$.

The following theorem tells us that $D_N(\gamma)$ is an informative metric to measure the dependence between the frequency and the severity components.

**Theorem 3.2.** *For $D_N(\gamma)$ in (3.11), the following are true:*

*(i) $D_N(\gamma)$ is well-defined if and only if $\gamma \le \log(1 + \tilde{r}_T/\nu_{T+1})$.*

*(ii) $D_N(\gamma)$ is a strictly increasing function of $\gamma$.*

*(iii)*

$$D_N(\gamma) = e^\gamma \left[1 - \left(\frac{\nu_{T+1}}{\tilde{r}_T}\right)(e^\gamma - 1)\right]^{-r_T - 1} \begin{cases} = 1 & \text{if} \quad \gamma = 0 \\ > 1 & \text{if} \quad \gamma > 0 \\ < 1 & \text{if} \quad \gamma < 0 \end{cases}. \tag{3.12}$$

*Proof.* (i) One can see that $D_N(\gamma)$ is well-defined if and only if $1 - \left(\frac{\nu_{T+1}}{\tilde{r}_T}\right)(e^\gamma - 1) \ge 0$, which is equivalent to $\gamma \le \log(1 + \tilde{r}_T/\nu_{T+1})$.

(ii) First, $g(\gamma) = 1 - \left(\frac{\nu_{T+1}}{\tilde{r}_T}\right)(e^\gamma - 1)$ is a strictly decreasing function of $\gamma$. Then one can see that $\left[1 - \left(\frac{\nu_{T+1}}{\tilde{r}_T}\right)(e^\gamma - 1)\right]^{-r_T - 1} = g(\gamma)^{-r_T - 1}$ is increasing function of $\gamma$ since $r_T > 0$. Finally, it is clear that $D_N(\gamma)$ is a product of two strictly increasing functions of $\gamma$, $g(\gamma)^{-r_T - 1}$ and $e^\gamma$.

(iii) It follows from (ii) and $D_N(0) = 1$.

$\square$

From (iii) of Theorem 3.2, if there is positive dependence, then the credibility premium of the compound sum would be greater than merely the product of the credibility premiums for each component. Therefore, one needs to compensate for the difference by multiplying an adjustment factor for dependence greater than one. In the case of negative dependence, the opposite holds so that the compensation for the difference will be an adjustment factor less than one. In addition, $D_N(\gamma)$ not only measures the sign of dependence but also the magnitude of dependence as we see in (ii) of Theorem 3.2. Finally, although the upper bound for a feasible value of $\gamma$ is $\log(1 + \tilde{r}_T/\nu_{T+1})$, it is not problematic in practice because usually $\nu_{T+1}$ is around 0.1 and $\tilde{r}_T \geq 3.1584$ in our model specification. Therefore, so long as $\gamma \leq 3.483821 = \ln(1 + 3.1584/0.1)$, $D_N(\gamma)$ is well-defined, which is satisfied with all observations in our subsequent empirical analysis.

One can also justify a relationship between $\gamma$ and the Pearson correlation coefficient. Since

$$\mathbb{E}\left[N_{T+1}|\mathbf{n}_T\right] \times \mathbb{E}\left[C_{T+1}|\mathbf{n}_T, \overline{\mathbf{c}}_T\right] = \left[1 - \left(\frac{\nu_{T+1}}{\tilde{r}_T}\right)(e^\gamma - 1)\right]^{-r_T} \times \mathbb{E}\left[\nu_{T+1}\theta^N|\mathbf{n}_T\right] \times \mathbb{E}\left[\tilde{\mu}_{T+1}\theta^C|\mathbf{n}_T, \overline{\mathbf{c}}_T\right]$$

and

$$\begin{aligned}\mathrm{Cov}\left[N_{T+1}, \overline{C}_{T+1}|\mathbf{n}_T, \overline{\mathbf{c}}_T\right] &= \mathbb{E}\left[S_{T+1}|\mathbf{n}_T, \overline{\mathbf{c}}_T\right] - \mathbb{E}\left[N_{T+1}|\mathbf{n}_T\right] \times \mathbb{E}\left[C_{T+1}|\mathbf{n}_T, \overline{\mathbf{c}}_T\right] \\ &= \left(e^\gamma \left[1 - \left(\frac{\nu_{T+1}}{\tilde{r}_T}\right)(e^\gamma - 1)\right]^{-1} - 1\right) \times \left[1 - \left(\frac{\nu_{T+1}}{\tilde{r}_T}\right)(e^\gamma - 1)\right]^{-r_T} \\ &\quad \times \mathbb{E}\left[\nu_{T+1}\theta^N|\mathbf{n}_T\right] \times \mathbb{E}\left[\tilde{\mu}_{T+1}\theta^C|\mathbf{n}_T, \overline{\mathbf{c}}_T\right]\end{aligned}$$

by (3.11), it can be shown that

$$\mathrm{Corr}\left[N_{T+1}, \overline{C}_{T+1}|\mathbf{n}_T, \overline{\mathbf{c}}_T\right] = \frac{e^\gamma \left[1 - \left(\frac{\nu_{T+1}}{\tilde{r}_T}\right)(e^\gamma - 1)\right]^{-1} - 1}{\mathrm{CV}_{N_{T+1}|\mathbf{n}_T} \cdot \mathrm{CV}_{\overline{C}_{T+1}|\mathbf{n}_T, \overline{\mathbf{c}}_T}},$$

where $\mathrm{CV}_X = \sqrt{\mathrm{Var}\left[X\right]}/\mathbb{E}\left[X\right]$, which implies that the sign of $\mathrm{Corr}\left[N_{T+1}, \overline{C}_{T+1}|\mathbf{n}_T, \bar{\mathbf{c}}_T\right]$ would be the same as that of $\gamma$. Further, as $\gamma$ becomes smaller (larger), $\mathrm{Corr}\left[N_{T+1}, \overline{C}_{T+1}|\mathbf{n}_T, \bar{\mathbf{c}}_T\right]$ also becomes smaller (larger). In this regard, $D_N(\gamma)$ can provide essentially the same information as the Pearson correlation coefficient between $N_{T+1}$ and $\overline{C}_{T+1}$ given $\mathbf{n}_T, \bar{\mathbf{c}}_T$, other than the fact that the Pearson correlation coefficient is calibrated to be between $-1$ and 1, whereas $D_N(\gamma)$ is not. The Pearson correlation coefficient also measures only linear relationships.

## 3.4 Data description

To calibrate the proposed models in this article, we use a dataset from a Singapore automobile insurance company, which contains both policy characteristics and claims experience. This dataset consists of observations of nine years, 1993-2001. The same dataset or sampled data has been used in other research articles, including but not limited to Frees and Valdez (2008) and Shi and Valdez (2012). In the dataset, the observations for the first eight years, 1993-2000, were used as training set for estimating the associated parameters in each model, whereas the observations for the last year, 2001, were used for out-of-sample validation. There are $M = 50,215$ unique policyholders who are tracked for $T_i$ years. It is clear that the maximum value of $T_i$ is 8 in this case.

The observed policy characteristics were used as covariates in both components, frequency and the average severity. Table 3.1 provides the summary statistics only for the training set. Descriptions of the covariates, which contain both driver and vehicle information, are also provided in this table. In total, nine variables were used as covariates, which can either be categorical or continuous. Gender and issue age are related to driver information, while the others are related to vehicle information. Although the table is self-explanatory, we provide a few remarks to explain the dataset more clearly. First, the proportion of female drivers in Singapore is generally quite lower than that in other developed countries. Second,

Table 3.1: Observable policy characteristics used as covariates in the training set

| Categorical variables | Description | | Proportions | |
|---|---|---|---|---|
| VehType | Type of insured vehicle: | Car | 99.27% | |
| | | MotorBike | 0.47% | |
| | | Others | 0.26% | |
| Gender | Insured's sex: | Male = 1 | 80.82% | |
| | | Female = 0 | 19.18% | |
| CoverCode | Type of insurance cover: | Comprehensive = 1 | 78.65% | |
| | | Others = 0 | 21.35% | |
| **Continuous variables** | | **Minimum** | **Mean** | **Maximum** |
| VehCapa | Insured vehicle's capacity in cc | 10.00 | 1587.44 | 9996.00 |
| VehAge | Age of vehicle in years | -1.00 | 6.71 | 48.00 |
| Age | The policyholder's issue age | 18.00 | 44.46 | 99.00 |
| NCD | No Claim Discount in % | 0.00 | 35.67 | 50.00 |

there are a few observations on insured motorbike but are not to be ignored in terms of risk characteristic. Finally, VehAge, the age of the vehicle in years, is defined as the difference between the issue year and the model year of the insured vehicle. Therefore, it is not unusual to have a -1 as an observed value of VehAge because it is possible to purchase a car in the year prior to the release of the model.

Table 3.2: Measures for assessing correlation between the frequency and the (log) average severity

| | Pearson | Kendall | Spearman |
|---|---|---|---|
| Estimate | 0.04052 | 0.04227 | 0.05201 |
| p-value | 0.00000 | 0.00000 | 0.00000 |

As a preliminary observation in understanding the relation between frequency and severity, we may consider the idea of correlation measure. Table 3.2 provides the values of estimated correlation measures, which include Pearson coefficient, Kendall's tau, and Spearman's rho. In all correlation measures, we see there is strong positive correlation between the frequency and (logarithm of) the average severity, which is also shown in Figure 3.1. Intuitively, this result is reasonable because we can think that the common risk characteristic of each policyholder affects both the frequency and the average severity. However, this might be only a preliminary

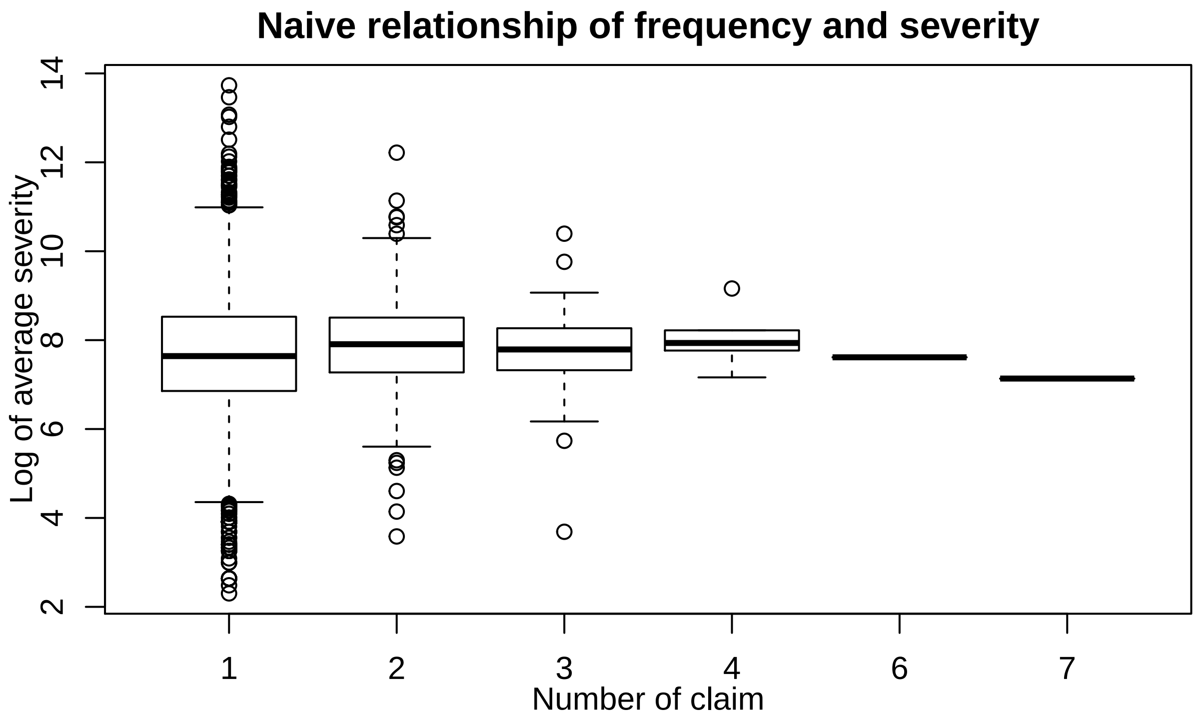


Figure 3.1: Naive relationship of frequency and severity

investigation since we have not controlled for any effects of observed characteristics of the policyholders which are usually heterogeneous.

It is well known that Poisson with gamma random effects leads to a negative binomial which usually fits better than a simple Poisson model, especially with the presence of overdispersion. In our training data, the overall average number of claims is 0.0941 whereas the variance is 0.1024 which indicate a possibility of overdispersion. This idea can be validated with a goodness-of-fit test for the Poisson in comparison to the negative binomial, which is shown in Table 3.3.

Table 3.3: Goodness-of-fit test for the frequency component

| Count | Observed | Poisson | Negative Binomial |
|---|---|---|---|
| 0 | 148198 | 147608.6 | 148193.5 |
| 1 | 12788 | 13895.4 | 12809.9 |
| 2 | 1109 | 654 | 1077.6 |
| 3 | 77 | 20.5 | 89.8 |
| 4+ | 7 | 0.5 | 8.1 |
| $\chi^2$ | | 647.4 | 2.9 |

Figure 3.2 provides log quantile-quantile (log-QQ) plots of fitting the gamma and generalized gamma distribution for the training set without consideration of the observed covariates. Since it can be difficult to decide whether G-gamma or gamma distribution shows better fit on the observed average severities with QQ plot, which is a qualitative and visual tool for model assessments, we need further analysis on the dataset that will consider both the effects of covariates and its longitudinal property.

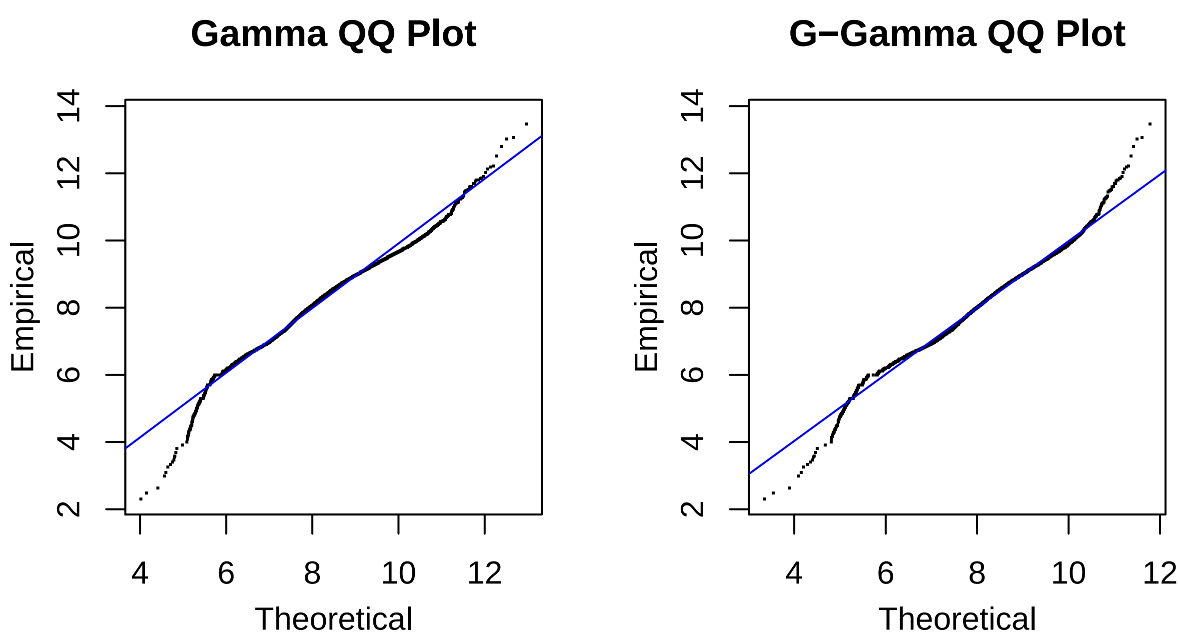


Figure 3.2: log-QQ plots of fitting gamma and G-gamma to average severity on training set

## 3.5 Results of estimation and model validation

Table 3.4 below provides the details of the estimation results from each of the frequency models described in section 3.2.1. In the estimation procedure, the hyperparameter $r$ in $\theta^N \sim \mathcal{G}(r, 1/r)$ was estimated from the joint likelihood of frequency component in (3.8), which ended up with $\hat{r} = 3.1584$. This value deems appropriate since 95% credible interval of $\theta^N$ would include $(0.54, 2.00)$ with $\hat{r} = 3.1584$; this has been proposed as the range of Belgian bonus-malus system based on claim frequency in Lemaire (1998). According to this table, the estimates from each model are more or less similar but MVNB model outperforms the naive Poisson model in terms of model selection criteria such as AIC and BIC.

Table 3.4: Regression estimates of the frequency models

| | Poisson | | | MVNB | | |
|---|---|---|---|---|---|---|
| | Estimate | s.e. | Pr(>\|t\|) | Estimate | s.e. | Pr(>\|t\|) |
| (Intercept) | -4.47 | 0.40 | 0.00 | -4.46 | 0.32 | 0.00 |
| VTypeCar | 0.19 | 0.20 | 0.34 | 0.09 | 0.12 | 0.45 |
| VTypeMBike | -1.49 | 0.54 | 0.01 | -1.64 | 0.55 | 0.00 |
| logVehCapa | 0.33 | 0.03 | 0.00 | 0.34 | 0.04 | 0.00 |
| VehAge | -0.02 | 0.00 | 0.00 | -0.01 | 0.00 | 0.00 |
| SexM | 0.11 | 0.02 | 0.00 | 0.12 | 0.02 | 0.00 |
| Comp | 0.96 | 0.04 | 0.00 | 1.00 | 0.04 | 0.00 |
| Age | -0.03 | 0.02 | 0.11 | -0.03 | 0.01 | 0.00 |
| Age $^2$ | 0.00 | 0.00 | 0.37 | 0.00 | 0.00 | 0.00 |
| Age $^3$ | 0.00 | 0.00 | 0.78 | 0.00 | 0.00 | 0.27 |
| NCD | -0.01 | 0.00 | 0.00 | -0.01 | 0.00 | 0.00 |
| Loglikelihood | | -48362.77 | | | -48230.29 | |
| AIC | | 96747.54 | | | 96484.58 | |
| BIC | | 96869.50 | | | 96604.54 | |

For the validation of the calibrated frequency models, we compare the actual claim count in the test set with the predicted claim count based on the corresponding observed covariates. As measures of prediction performance, root-mean-square error (RMSE) and mean absolute error (MAE) are considered. RMSE and MAE measure the discrepancy between the actual loss and predicted loss in terms of $L_2$ and $L_1$ norms, respectively. Even though there are no substantital differences in all validation measures as shown in Table 3.5, the use of frequency random effects model can still be worthy as a way of incentivizing policyholders to have fewer accidents so that they can get discounts on their premium in subsequent years.

Table 3.5: Validation measures for the frequency models

| | Poisson | MVNB |
|---|---|---|
| RMSE | 0.31071 | 0.31074 |
| MAE | 0.15455 | 0.15389 |

Table 3.6 below provides the details of the estimation results from each of the average severity models described in section 3.2.2. In both calibration of MVGP and MVGB2 models,

the hyperparameter $k$ was estimated from the joint likelihoods of MVGP and MVGB2 distributions in (3.9) and (3.10), respectively. In the case of MVGP model, $\hat{k} = 17.0227$ so that when $\theta^C \sim \mathcal{IG}(k+1, k)$, 95% credible interval of $\theta^C$ can be around $(0.59, 1.50)$, which is narrower than that of $\theta^N$. This can be validated because claim severity is not used for bonus-malus systems for many countries except for South Korea, which supports the assertion that there is less variability on the severity component random effects. According to Table 3.6, all heavy-tailed models (GB2, MVGP, and MVGB2) outperform the Gamma GLMM in terms of our usual model selection criteria, AIC and BIC. Further, with the given empirical data, AIC of the naive gamma model is 277236.83, which means there are substantial improvement in model selection criteria with the aforementioned models. Therefore, one can claim that the proposed models fit the empirical data in the training set well compared to the naive gamma model, which is an industry benchmark.

Note that except for Gamma GLMM, the sign of the coefficient of claim count as a covariate for the average severity is significantly negative, which implies strong negative correlation between frequency and the average severity component. Such negative dependence has been observed in other lines of insurance as in Frees et al. (2011a).

Table 3.6: Regression estimates of the average severity models

| | GB2 | | Gamma GLMM | | MVGP | | MVGB2 | |
|---|---|---|---|---|---|---|---|---|
| | Estimate | Pr(>\|t\|) | Estimate | Pr(>\|t\|) | Estimate | Pr(>\|t\|) | Estimate | Pr(>\|t\|) |
| (Intercept) | 7.79 | 0.00 | 6.43 | 0.00 | 8.07 | 0.00 | 8.11 | 0.00 |
| VTypeCar | -1.55 | 0.00 | 0.12 | 0.62 | 1.17 | 0.00 | 1.05 | 0.00 |
| VTypeMBike | 3.94 | 0.00 | 2.32 | 0.00 | 2.24 | 0.00 | 3.42 | 0.00 |
| logVehCapa | 0.66 | 0.00 | 0.33 | 0.00 | 0.24 | 0.00 | 0.25 | 0.00 |
| VehAge | 0.07 | 0.00 | -0.01 | 0.00 | -0.01 | 0.00 | -0.02 | 0.00 |
| SexM | -0.07 | 0.01 | -0.02 | 0.49 | 0.05 | 0.04 | -0.04 | 0.09 |
| Comp | 1.38 | 0.00 | 0.19 | 0.00 | 0.12 | 0.01 | 0.14 | 0.00 |
| Age | -0.21 | 0.00 | -0.05 | 0.03 | -0.16 | 0.00 | -0.15 | 0.00 |
| Age $^2$ | 0.00 | 0.00 | 0.00 | 0.02 | 0.00 | 0.00 | 0.00 | 0.00 |
| Age $^3$ | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 |
| NCD | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Count | -0.13 | 0.00 | 0.01 | 0.65 | -0.19 | 0.00 | -0.07 | 0.00 |
| Loglikelihood | | -126052.00 | | -133760.00 | | -125882.00 | | -125119.00 |
| AIC | | 252134.37 | | 267548.24 | | 251792.95 | | 250268.32 |
| BIC | | 252247.07 | | 267653.42 | | 251898.14 | | 250381.01 |

Table 3.7: Summary of adjustment factors for the average severity models

| | GB2 | Gamma GLMM | MVGP | MVGB2 |
|---|---|---|---|---|
| Min | 0.78 | 1.01 | 0.70 | 0.87 |
| Max | 0.88 | 1.02 | 0.83 | 0.93 |
| Mean | 0.87 | 1.01 | 0.82 | 0.92 |
| Range | 0.10 | 0.01 | 0.13 | 0.06 |
| Std. Dev | 0.01 | 0.00 | 0.01 | 0.00 |

As shown in (3.12), negative (positive) value of $\hat{\gamma}$ implies negative (positive) correlation between the frequency and severity. As a result, we observe that adjustment factors for GB2, MVGP, MVGB2 models are all below 1, which accounts for negative dependence between the frequency and severity. On the other hand, the Gamma GLMM has an adjustment factor that slightly larger than 1, although not statistically significant. Summary for the values of $D_N(\gamma)$, the dependence adjustment factor, under these models is provided in Table 3.7.

For a posterior ratemaking framework, one can analyze how past claim records on claim frequency and/or severity affect the credibility factors, which are multiplied by the prior premium according to observed characteristics. Figure 3.3 demonstrates the distribution of credibility factors based on claims history, where number of claims observed means $\sum_{t=1}^{T} N_t$ and average severity observed means $\frac{\sum_{t=1}^{T} S_t}{\sum_{t=1}^{T} N_t}$, respectively. We observe a natural pattern that as the number of previous claims increases, credibility factor on the frequency component rises which incurs surcharge on the pure premium. We also see that if the historical average severity is relatively small, then credibility factor for the severity component might be slightly less than 100%. However, it does not mean that the policyholder would enjoy discount on total pure premium since given any previous claim, it is expected that credibility factor for the frequency component would be greater than 100% in general. On the contrary, if $\sum_{t=1}^{T} N_t = 0$, then credibility factor for the severity component would be exactly 100% so that the policyholder can fully enjoy premium discount due to credibility factor on the frequency, which would be lower than 100%. Such examples of possible credibility weights depending on claim experiences are further demonstrated in Table 3.8. More in-depth applications of our dependent compound risk model can be done in several directions such as using other risk measures, like conditional tail expectations, to better summarize risk distributions, as well as analyzing coverage modifications and reinsurance treaties. We refer the readers to Frees et al. (2009) for such applications.

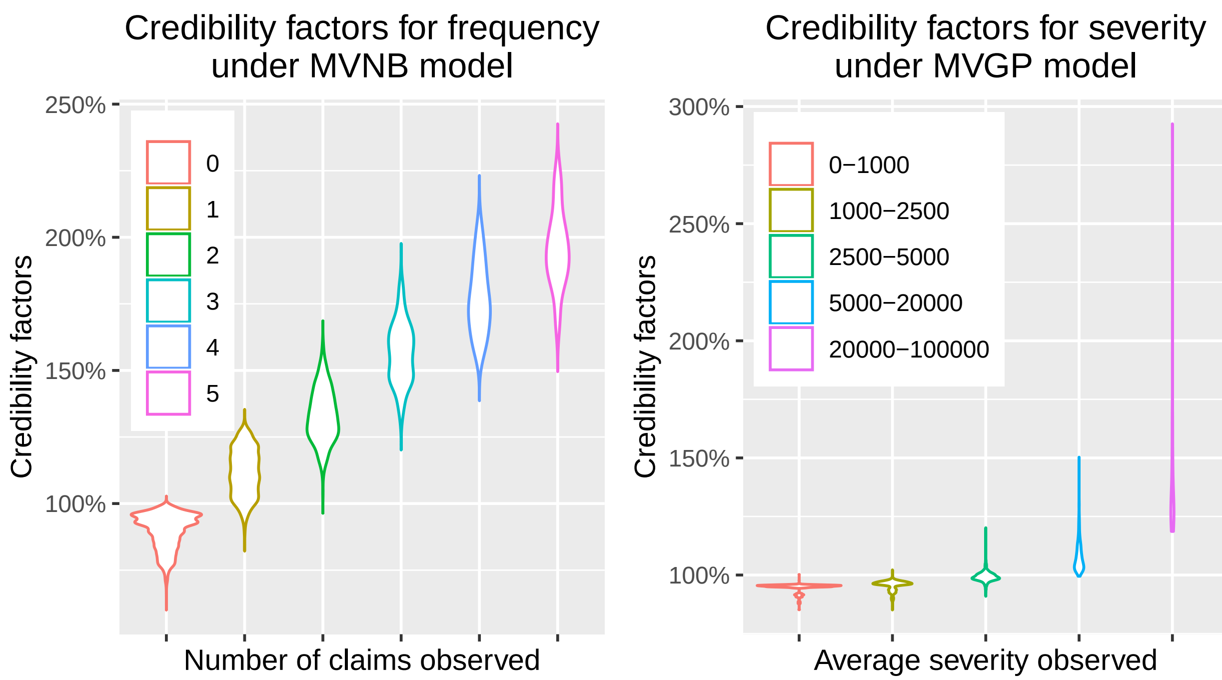



Figure 3.3: Credibility factors for frequency and severity components

Table 3.8: Credibility factors of selected policyholders under MVNB and MVGP models

| | Observed claims | | Credibility factors | | |
|---|---|---|---|---|---|
| ID | Counts | Total losses | Frequency | Severity | Total |
| 1614 | 0 | 0 | 83% | 100% | 83% |
| 371 | 2 | 1173 | 154% | 92% | 141% |
| 1816 | 1 | 32250 | 114% | 148% | 169% |

For the validation of the calibrated average severity models, first we compare the actual compound sum of the claims in the test set with the predicted compound sum of the claims, which are calculated using the formula for the credibility premium of compound sum in (3.11). From Table 3.9, we see that MVGP model outperforms the other models in terms of RMSE whereas GB2 model is the best in terms of MAE validation measure.

Table 3.9: Validation measures for the average severity models

| | GB2 | Gamma GLMM | MVGP | MVGB2 |
|---|---|---|---|---|
| RMSE | 2504.3 | 2486.5 | 2484.9 | 2500.0 |
| MAE | 524.0 | 569.7 | 549.7 | 567.9 |

It can be of interest to insurance companies not only to obtain individual expected claims, but also the predictive distribution of future claims from an entire portfolio for better enterprise risk management. To further visualize the quality of the different models, we compare the empirical distribution of the actual compound sum of the whole portfolio from the validation set, with the predictive distributions under each average severity model as follows:

1. Generate $N_{i,2001}^{(r)}$ from the fitted MVNB model, using the predictive distribution of $N_{i,2001}|N_{i,1993},\ldots,N_{i,2000}$ estimated from MVNB model and available covariate information $\mathbf{x}_{i,2001}$ for $i=1,\ldots,M$ and $r=1,\ldots,100$.

2. Provided $N_{i,2001}^{(r)}>0$, generate $\overline{C}_{i,2001}^{(r)}>0$ given $\mathbf{x}_{i,2001}$ and $N_{i,2001}^{(r)}>0$ under each average severity model. For example, one can easily generate a GB2 random sample from given predictive distribution because $Y := c\cdot\sqrt[p]{\frac{U}{1-U}} \sim \mathcal{GB}2(k+1,c,\psi,p)$ where $U\sim \text{Beta}(\psi,k+1)$. Now, $r^{th}$ simulated sample of compound sum for policyholder $i$ is given as follows:

$$S_{i,2001}^{(r)} = \begin{cases} N_{i,2001}^{(r)}\cdot\overline{C}_{i,2001}^{(r)}, & N_{i,2001}^{(r)}>0 \\ 0, & N_{i,2001}^{(r)}=0 \end{cases},$$

   whereas $S_{i,2001}$ denotes actual value of compound sum for policyholder $i$ in the validation set. Note that since we are comparing predictive distribution of $S$ under each severity model, we only utilize the data points where $S_{i,2001}>0$ and $S_{i,2001}^{(r)}>0$ for $i=1,\ldots,M$ and $r=1,\ldots,100$.

3. Compute kernel density of actual compound sums of whole portfolio from validation set using

$$\{\log(S_{i,2001}) \mid S_{i,2001}>0,\ i=1,\ldots,M\}.$$

Likewise, kernel density of predicted compound sums of whole portfolio is computed using

$$\{\log(S_{i,2001}^{(r)}) \mid S_{i,2001}^{(r)} > 0,\, i = 1, \ldots, M,\, r = 1, \ldots, 100\}$$

under each severity model.

According to Figure 3.4 and Tables 3.10 and 3.11, the results show that Gamma GLMM fail to describe the overall distribution of the compound sum $S$ for the whole portfolio. It is noteworthy to observe that Gamma GLMM still remains to be a light-tailed distribution and it fails to describe the overall distribution of $S$ appropriately although it outperforms GB2 and MVGB2 models in terms of RMSE, which is similar result to that of Jeong et al. (2020). On the other hand, heavy-tailed distributions, GB2, MVGP and MVGB2 models, are quite close to the kernel density of the actual losses from whole portfolio and well describe overall risks distribution of the portfolio. Note that the claim data were collected at the end of calender year 2003 so that some of the claims in calender year 2001 might not have yet been fully reported or developed, which explains the left-skewed shape of the distribution of the observed $S$.

Table 3.10: Summary statistics for predicted log of compound sum

| | Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. | St.dev |
|---|---|---|---|---|---|---|---|
| Actual | 3.7028 | 6.5004 | 7.5171 | 7.4981 | 8.4012 | 12.4315 | 1.1585 |
| GB2 | -16.5643 | 6.7629 | 7.7077 | 7.5763 | 8.5373 | 13.3234 | 1.4082 |
| Gamma GLMM | -65.8374 | 1.5173 | 5.7241 | 3.9148 | 8.1959 | 13.4303 | 6.2099 |
| MVGP | -9.0089 | 7.1225 | 8.0047 | 7.8051 | 8.7178 | 12.2484 | 1.3071 |
| MVGB2 | -0.2072 | 7.1611 | 8.0029 | 7.9502 | 8.8001 | 14.5319 | 1.2912 |

Table 3.11: Summary statistics for predicted compound sum

| | Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. | St.dev |
|---|---|---|---|---|---|---|---|
| Actual | 40.6 | 665.4 | 1839.2 | 3572.0 | 4452.5 | 250571.0 | 8498.2 |
| GB2 | 0.0 | 865.1 | 2225.4 | 4247.6 | 5101.7 | 611356.0 | 6846.7 |
| Gamma GLMM | 0.0 | 4.6 | 306.2 | 4701.0 | 3626.1 | 680316.4 | 11809.4 |
| MVGP | 0.0 | 1239.6 | 2995.0 | 4516.6 | 6110.8 | 208656.0 | 4943.2 |
| MVGB2 | 0.8 | 1288.4 | 2989.5 | 6308.1 | 6635.1 | 2047047.0 | 17190.6 |

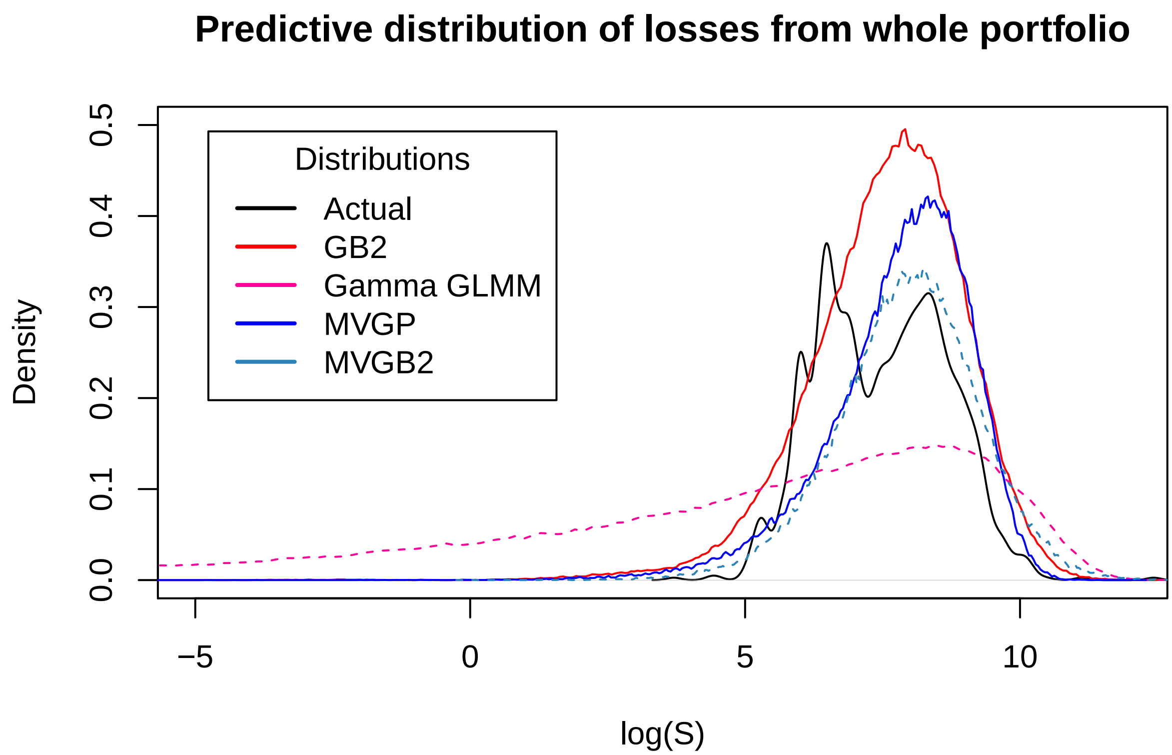



Figure 3.4: Predictive distribution of compound sums from whole portfolio under severity models

Finally, even though GB2, MVGP, and MVGB2 models can capture the heavy-tailed behavior of actual claims of the overall portfolio, we expect a predictive model should be able to do risk classification appropriately, so that those who were charged with higher premium are expected to incur higher loss amounts. In that sense, we compare Gini index proposed in Frees et al. (2014b) to measure risk classification performance of GB2, MVGP, and MVGB2 models. We prefer models with higher Gini indices in terms of risk classification performance. For detailed calculations and statistical properties of the Gini index, see Frees et al. (2011b). Figure 3.5 shows us that MVGP model has the highest Gini index while GB2 model has the lowest Gini index. Thus, we can conclude that MVGP model may be favored by not only its flexibility to describe heavy-tail behaviors, but also its appropriateness for classifying risks utilizing individual claims history for credibility premiums.

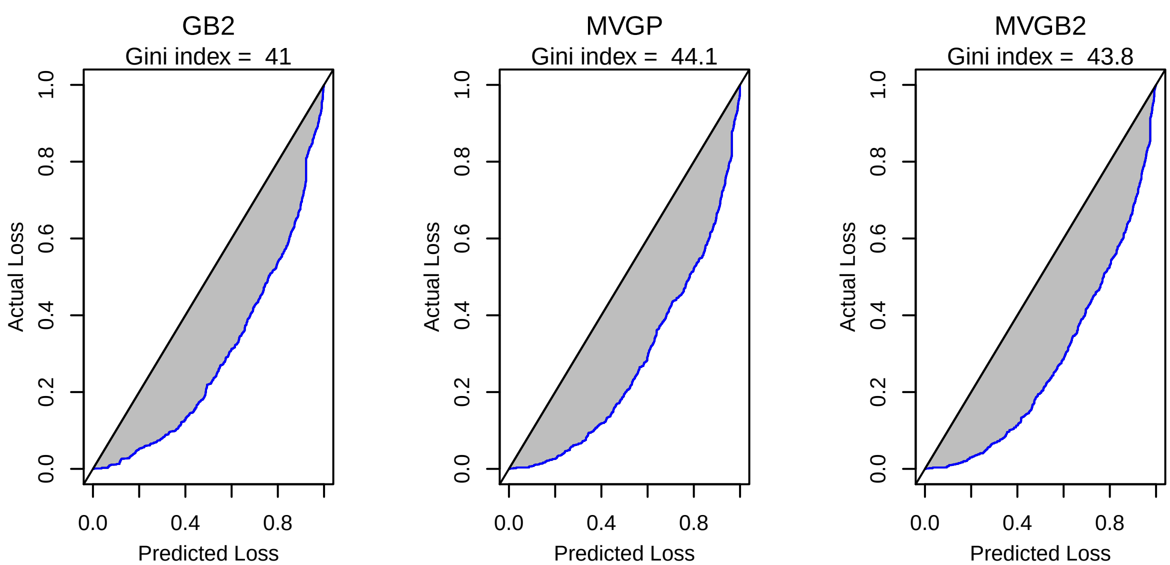



Figure 3.5: Gini indices for severity models

## 3.6 Conclusion

In actuarial practice, both the longitudinal property of general insurance data and the possible dependence between the claim frequency and the average severity cannot be overlooked for the construction of a ratemaking model. In this article, we explore the possiblity of incorporating these two relevant aspects in a ratemaking model via the use of dependent compound random effects models. These random effects models are based on the conjugate family of distributions in both frequency and severity. These models allowed us to derive credibility premiums for the compound sum incorporating a dependence function that is informative as a measure of the strength and direction of the association between frequency and severity. In our model calibration based on empirical claims from a Singapore auto insurance company, the results show us that proposed MVGP and MVGB2 distributions outperform naive Gamma GLM and Gamma GLMM in terms of improved predictions and analytical tractability. We further demonstrated how these models can be used to analyze the predictive distributions of total or aggregate sum of claims and to evaluate or understand the reasonableness of credibility factors.

# Chapter 4

# Bayesian credibility premium with GB2 copulas

## 4.1 Introduction

[1] For our purpose, a random variable is always well defined on a given probability space and all random variables are continuous. Consider a sequence of random variables $Y_1, \ldots, Y_T$ with the following Bayesian framework:

- $Y_t|\Theta \sim f_{Y_t}(y_t|\theta)$ are independent for $t = 1, \ldots, T$ and
- $\Theta \sim p(\theta)$ is the prior distribution.

It is easy to see that the posterior density of $\Theta|\mathbf{Y}_T$ has the expression

$$p(\theta|\mathbf{Y}_T = \mathbf{y}_T) = \frac{\prod_{t=1}^{T} f_{Y_t}(y_t|\theta)p(\theta)}{\int \prod_{t=1}^{T} f_{Y_t}(y_t|\theta)p(\theta)d\theta} = \frac{\prod_{t=1}^{T} f_{Y_t}(y_t|\theta)p(\theta)}{h_T(\mathbf{y}_T)} \tag{4.1}$$

[1]Most part of this chapter is from Jeong and Valdez (2020a).

and its posterior mean is given by $\mathbb{E}\left[\Theta|\mathbf{Y}_T=\mathbf{y}_T\right]=\int \theta\, p(\theta|\mathbf{y}_T)d\theta$, where $\mathbf{Y}_T=(Y_1,\ldots,Y_T)'$, $\mathbf{y}_T=(y_1,\ldots,y_T)'$, and $h_T(\mathbf{y}_T)$ is the multivariate density function for $\mathbf{Y}_T$.

Define a new observation for $t=T+1$ as $Y_{T+1}$ whereby $Y_{T+1}|\Theta$ is independent with $Y_t|\Theta$ for $t=1,\ldots,T$. Then we can define the expectation of the new observation, given $\Theta$, as

$$\mu(\Theta)=\mathbb{E}\left[Y_{T+1}|\Theta\right]=\int y_{T+1}f_{Y_{T+1}}(y_{T+1}|\theta)dy_{T+1} \tag{4.2}$$

and the Bayesian posterior mean as

$$\mathbb{E}\left[\mu(\Theta)|\mathbf{Y}_T=\mathbf{y}_T\right]=\int \mu(\theta)\, p(\theta|\mathbf{y}_T)d\theta$$

in the same manner as Bühlmann and Gisler (2006). We obtain the following lemma.

**Lemma 4.1.** *Given the above Bayesian specification, the following relation holds:*

$$\mathbb{E}\left[\mu(\Theta)|\mathbf{Y}_T=\mathbf{y}_T\right]=\mathbb{E}\left[Y_{T+1}|\mathbf{Y}_T=\mathbf{y}_T\right]. \tag{4.3}$$

*Proof.* From equations (4.1) and (4.2), we have

$$\begin{aligned}
\mathbb{E}\left[\mu(\Theta)|\mathbf{Y}_T=\mathbf{y}_T\right] &= \int\left(\int y_{T+1}f_{Y_{T+1}}(y_{T+1}|\theta)dy_{T+1}\right)\prod_{t=1}^{T} f_{Y_t}(y_t|\theta)p(\theta)\Big/h_T(\mathbf{y}_T)\,d\theta \\
&= \int\left(\int \frac{\prod_{t=1}^{T+1} f_{Y_t}(y_t|\theta)p(\theta)}{h_{T+1}(\mathbf{y}_{T+1})}d\theta\right) y_{T+1}\frac{h_{T+1}(\mathbf{y}_{T+1})}{h_T(\mathbf{y}_T)}dy_{T+1} \\
&= \int y_{T+1}f_{Y_{T+1}|Y_{\mathbf{T}}}(y_{T+1}|\mathbf{y}_T)dy_{T+1}=\mathbb{E}\left[Y_{T+1}|\mathbf{Y}_T=\mathbf{y}_T\right].
\end{aligned}$$

$\square$

As a special case, let us assume the following model specification:

$$Y_1,\ldots,Y_T,Y_{T+1}|\Theta \overset{i.i.d.}{\sim} N(\Theta,(1-\rho)\sigma^2) \quad \text{and} \quad \Theta \;\sim\; N(\mu,\rho\,\sigma^2).$$

Then $\mathbf{Y}_T$ has a multivariate normal distribution with a common pairwise correlation of $\rho$ so that we can write the correlation matrix as

$$\boldsymbol{\Sigma}_T = \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{pmatrix} = (1-\rho)\left(\mathbf{I}_T + \frac{\rho}{1-\rho}\mathbf{1}_T\mathbf{1}_T'\right),$$

where $\mathbf{1}_T' = (1, 1, \ldots, 1)$ is a row vector of 1's with dimension $T$ and $\mathbf{I}_T$ is a $T \times T$ identity matrix. In this case, we can show that all the marginals are also identically distributed as normal with mean $\mu$ and variance $\sigma^2$ and the Bayesian posterior mean has the familiar form of a credibility premium, as shown in Berger (1985):

$$\mathbb{E}\left[Y_{T+1} | \mathbf{Y}_T = \mathbf{y}_T\right] = \frac{\rho T}{1-\rho+\rho T} \cdot \overline{Y} + \frac{1-\rho}{1-\rho+\rho T} \cdot \mu \tag{4.4}$$

where $\overline{Y} = \sum_{t=1}^{T} y_t \Big/ T$, the average of past values.

The Bayesian specification captures the heterogeneity of each subject in a longitudinal dataset given observed values of $y_1, \ldots y_T$. The Bayesian posterior mean in equation (4.3) is the predicted value for the next time period $T+1$ given the historical observations. In the context of insurance, equation (4.4) is called a credibility-weighted premium and is used to calculate the subsequent year's premium given the past history of claims, for claims that follow the normal distribution.

This paper extends this result in the framework where we have a multivariate GB2. In Section 4.2, we show a construction of the multivariate GB2 and study its properties. In Section 4.3, we briefly define the concept of copula and introduce the Bayesian credibility premium as an expectation under a change of probability measure. In Section 4.4, we show that using this change of measure, it becomes straightforward to derive Bayesian credibility premium with

GB2 copulas. We consider a numerical illustration in Section 4.5. Section 4.6 discusses the generalized Pareto as a special case. Finally, we provide concluding remarks in Section 4.7.

## 4.2 The multivariate GB2 distribution

To construct the multivariate GB2 distribution, consider the following model specification:

$$Y_t|\Theta \overset{i.i.d.}{\sim} \text{G-Gamma}\left(\psi, \frac{\Gamma(\psi)}{\Gamma(\psi+1/p)}\Theta, p\right) \quad \text{and} \quad \Theta \sim \text{GI-Gamma}\left(k, \frac{\Gamma(k)}{\Gamma(k-1/p)}\mu, p\right)$$

where

- G-Gamma $\left(\psi, \frac{\Gamma(\psi)}{\Gamma(\psi+1/p)}\Theta, p\right)$ refers to a generalized gamma with density

$$f_{Y_t}(y|\theta) = \frac{p}{y\Gamma(\psi)}\left(\frac{y\Gamma(\psi+1/p)}{\Gamma(\psi)\theta}\right)^{p\psi} \exp\left[-\left(\frac{y\Gamma(\psi+1/p)}{\Gamma(\psi)\theta}\right)^{p}\right]$$

  and

- GI-Gamma $\left(k, \frac{\Gamma(k)}{\Gamma(k-1/p)}\mu, p\right)$ refers to a generalized inverse-gamma with density

$$p(\theta) = \frac{p}{\theta\Gamma(k)}\left(\frac{\Gamma(k)\mu}{\Gamma(k-1/p)\theta}\right)^{pk} \exp\left[-\left(\frac{\Gamma(k)\mu}{\Gamma(k-1/p)\theta}\right)^{p}\right],$$

for $p > 0$. This specification can easily lead us to the following multivariate GB2 distribution by integrating out the scale parameter $\Theta$:

$$h_T(\mathbf{y}_T) = \int_0^\infty \prod_{t=1}^{T} f_{Y_t}(y_t|\theta)p(\theta)d\theta = \frac{p^T}{B(\Psi_T)\prod_{t=1}^{T} y_t} \frac{c^{pk}\prod_{t=1}^{T} y_t^{p\psi}}{\left[\sum_{t=1}^{T} y_t^p + c^p\right]^{\psi T+k}}, \tag{4.5}$$

where $B(\Psi_T)$ is the multivariate beta function, $c = \mu\frac{\Gamma(k)}{\Gamma(k-1/p)}\frac{\Gamma(\psi)}{\Gamma(\psi+1/p)}$, and $\Psi_T =$

$(\psi, \psi, \ldots, \psi, k)$ is a vector of size $(T+1)$ with equal first $T$ elements. Since $Y_t|\Theta$ are i.i.d., then the marginal distribution is straightforward to determine by setting $T = 1$ so that

$$f_t(y_t) = \frac{p}{y_t\, B(\psi, k)} \frac{c^{pk}\, y_t^{p\psi}}{\left[y_t^p + c^p\right]^{\psi+k}}, \tag{4.6}$$

where $B(\Psi_1) = B(\psi, k)$ with $B(\cdot, \cdot)$ is the beta function. For a random variable $Y_t$ with density given in equation (4.6), we can write $Y_t \sim \text{GB2}(k, c, \psi, p)$. See McDonald and Butler (1990) and McDonald and Bookstaber (1991) for applications of univariate GB2. A similar derivation of the multivariate GB2 has appeared in Yang et al. (2011), but the parameterization is different from above.

From the model specification, we can easily deduce the following moment properties:

$$\mathbb{E}\left[Y_t|\Theta\right] = \Theta \qquad \text{and} \qquad \text{Cov}(Y_t, Y_{t'}|\Theta) = \begin{cases} \Theta^2 \left[\dfrac{B(\psi+2/p, \psi)}{B(\psi+1/p, \psi+1/p)} - 1\right], & \text{for } t = t' \\ 0, & \text{for } t \neq t' \end{cases}$$

and

$$\mathbb{E}\left[\Theta\right] = \mu, \qquad \text{and} \qquad \text{V}ar\left[\Theta\right] = \mu^2 \left(\frac{B(k-2/p, k)}{B(k-1/p, k-1/p)} - 1\right).$$

These properties lead us to the unconditional mean and variance, respectively, of a GB2 distribution:

$$\begin{aligned} \mathbb{E}\left[Y_t\right] &= \mathbb{E}\left[\mathbb{E}\left[Y_t|\Theta\right]\right] = \mathbb{E}\left[\Theta\right] = \mu, \\ \text{V}ar\left[Y_t\right] &= \mathbb{E}\left[\text{V}ar\left[Y_t|\Theta\right]\right] + \text{V}ar\left[\mathbb{E}\left[Y_t|\Theta\right]\right] = \mathbb{E}\left[\Theta^2\left(\frac{B(\psi+2/p, \psi)}{B(\psi+1/p, \psi+1/p)} - 1\right)\right] + \text{V}ar\left[\Theta\right] \\ &= \mu^2 \left[\frac{B(k-2/p, k)}{B(k-1/p, k-1/p)} \frac{B(\psi+2/p, \psi)}{B(\psi+1/p, \psi+1/p)} - 1\right]. \end{aligned} \tag{4.7}$$

The following lemma shows that the multivariate GB2 has a pairwise correlation structure.

**Lemma 4.2.** *Suppose* $(Y_1, \ldots, Y_T)'$ *follows a multivariate GB2 distribution as given in* (4.5).

*Then they have a so-called pairwise correlation structure. In other words,*

$$Corr(Y_t, Y_{t'}) = \begin{cases} 1, & if \;\; t = t' \\ \\ \rho_{p,k,\psi}, & if \;\; t \neq t' \end{cases} \qquad where \;\; \rho_{p,k,\psi} = \frac{\frac{B(k-2/p,k)}{B(k-1/p,k-1/p)} - 1}{\frac{B(k-2/p,k)}{B(k-1/p,k-1/p)} \frac{B(\psi+2/p,\psi)}{B(\psi+1/p,\psi+1/p)} - 1}.$$

*Proof.* By definition, $\mathrm{Corr}(Y_t, Y_t) = 1$. It is straightforward to see that in the case where $t \neq t'$, we have

$$\begin{aligned} \mathrm{Cov}(Y_t, Y_{t'}) &= \mathbb{E}\left[\mathrm{Cov}(Y_t, Y_{t'}|\Theta)\right] + \mathrm{Cov}(\mathbb{E}\left[Y_t|\Theta\right], \mathbb{E}\left[Y_{t'}|\Theta\right]) \\ &= \mathrm{Cov}(\Theta, \Theta) = \mu^2 \left[\frac{B(k-2/p,k)}{B(k-1/p,k-1/p)} - 1\right] \end{aligned}$$

and from equation (4.7),

$$\mathrm{V}ar\left[Y_t\right] = \mathrm{V}ar\left[Y_{t'}\right] = \mu^2 \left[\frac{B(k-2/p,k)}{B(k-1/p,k-1/p)} \frac{B(\psi+2/p,\psi)}{B(\psi+1/p,\psi+1/p)} - 1\right].$$

The result for $t \neq t'$ immediately follows because $\mathrm{Corr}(Y_t, Y_{t'}) = \dfrac{\mathrm{Cov}(Y_t, Y_{t'})}{\sqrt{\mathrm{V}ar\left[Y_t\right] \mathrm{V}ar\left[Y_{t'}\right]}} = \dfrac{\mathrm{Cov}(Y_t, Y_{t'})}{\mathrm{V}ar\left[Y_t\right]}$. $\square$

Since it is known that for any $\alpha$, $\lim_{k\to\infty} \dfrac{\Gamma(k+\alpha)}{\Gamma(k)k^{\alpha}} = 1$, we have that for any fixed $p$,

$$\begin{aligned} \lim_{k\to\infty} \frac{B(k-2/p,k)}{B(k-1/p,k-1/p)} &= \lim_{k\to\infty} \frac{\Gamma(k-2/p)\Gamma(k)}{\Gamma(k-1/p)^2} \\ &= \lim_{k\to\infty} \frac{\Gamma(k)}{\Gamma(k-1/p)(k)^{1/p}} \frac{\Gamma(k-2/p)(k-1/p)^{1/p}}{\Gamma(k-1/p)} \frac{k^{1/p}}{(k-1/p)^{1/p}} = 1. \end{aligned}$$

This implies $\lim_{k\to\infty} \mathrm{Cov}(Y_t, Y_{t'}) = \lim_{k\to\infty} \rho_{p,k,\psi} = 0$ if $t \neq t'$.

The following lemma shows that conditional distribution of each component of a multivariate GB2 random vector remains a member of the GB2 family.

**Lemma 4.3.** *Suppose* $(Y_1, \ldots, Y_T)'$ *follows a multivariate GB2 distribution as given in* (4.5).

*Then for any* $j = 1, \ldots, T$,

$$(Y_j | Y_1, \ldots, Y_{j-1}, Y_{j+1}, \ldots, Y_T) \sim GB2(k + \psi(T-1), c^*, \psi, p),$$

*where* $c^* = c\left(1 + \sum_{t \neq j} (y_t/c)^p\right)^{1/p}$.

*Proof.* The proof is straightforward by noting that

$$f(Y_j | Y_1, \ldots, Y_{j-1}, Y_{j+1}, \ldots, Y_T) = \frac{h_T(\mathbf{y}_T)}{\int_0^\infty h_T(\mathbf{y}_T) dy_j}.$$

$\square$

## 4.3 Change of probability measure with copulas

Copula is a widely used method to model dependency among multivariate observations. It has increased in popularity in recent years because of its widespread applications in several disciplines including, but not limited to, medical science, demography, hydrology, insurance, finance, and engineering. With copulas, one can decompose the marginal distributions and their dependence structure. See Li (2000), Frees and Valdez (1998), Hougaard et al. (1992), Shih and Louis (1995), and Salvadori and De Michele (2007).

Copulas are functions that join (or couple) the multivariate distribution functions to their one-dimensional marginal distributions functions. See Joe (1997). Specifically, we have

$$H_T(y_1, \ldots, y_T) = \mathbb{P}(Y_1 \leq y_1, \ldots, Y_T \leq y_T) = C_T(F_1(y_1), \ldots, F_T(y_T))$$

where $F_t(\cdot)$ refers to the marginal distribution associated with $Y_t$, $H_T$ is their joint distribution function, and $C_T$ is the corresponding copula function where subscripts of $H$ and $C$ refer to the dimension of the random vector. Sklar (1959) proved the existence of copulas for every joint distribution function and demonstrated that they are indeed unique if the marginal distribution functions are continuous. It is also sometimes convenient to write this as

$$H_T(y_1, \ldots, y_T) = C_T(u_1, \ldots, u_T) \tag{4.8}$$

where $\mathbf{u}_T = (u_1, \ldots, u_T)'$ and $u_t = F_t(y_t)$ with $F_t$ the marginal distribution for $t = 1, \ldots, T$. Vectors shall be written in bold letters. For example, we shall denote the observed values of $\mathbf{Y}_T$ by $\mathbf{y}_T = (y_1, \ldots, y_T)'$ and similarly for $\mathbf{Y}_{T+1}$ by $\mathbf{y}_{T+1} = (y_1, \ldots, y_T, y_{T+1})'$. For ease of notation, we will also denote the vectors $\mathbf{u}_T = (u_1, \ldots, u_T)'$ and $\mathbf{u}_{T+1} = (u_1, \ldots, u_T, u_{T+1})'$. We shall assume that the densities of the copulas exist and are respectively denoted by

$$c_T(\mathbf{u}_T) = \frac{\partial^T C_T(\mathbf{u}_T)}{\partial u_1 \ldots \partial u_T}, \tag{4.9}$$

and

$$c_{T+1}(\mathbf{u}_{T+1}) = \frac{\partial^{T+1} C_{T+1}(\mathbf{u}_{T+1})}{\partial u_1 \ldots \partial u_T \partial u_{T+1}}. \tag{4.10}$$

Notice that the marginal distribution functions have been denoted by

$$F_t(y_t) = \mathbb{P}(Y_t \leq y_t) \quad \text{for} \quad t = 1, \ldots, T, T+1,$$

and if the corresponding density functions exist, we denote them by

$$f_t(y_t) = \frac{dF_t(y_t)}{dy_t} \quad \text{for} \quad t = 1, \ldots, T, T+1.$$

Similarly, the multivariate density functions, if they exist, will be respectively denoted by

$$h_T(\mathbf{y}_T) = \frac{\partial^T H_T(\mathbf{y}_T)}{\partial y_1 \ldots \partial y_T} \qquad \text{and} \qquad h_{T+1}(\mathbf{y}_{T+1}) = \frac{\partial^{T+1} H_{T+1}(\mathbf{y}_{T+1})}{\partial y_1 \ldots \partial y_T \partial y_{T+1}}.$$

In the following theorem, we derive the fundamental building block for deriving the Bayesian credibility premium, $\mathbb{E}\left[Y_{T+1}|\mathbf{Y}_T = \mathbf{y}_T\right]$, within a copula framework.

**Theorem 4.1.** *Consider the copula model satisfying the assumptions described in this section. The conditional expectation of $Y_{T+1}|\mathbf{Y}_T$ can be expressed in the following manner:*

$$\mathbb{E}\left[Y_{T+1}|\mathbf{Y}_T = \mathbf{y}_T\right] = \int y_{T+1} \frac{c_{T+1}(\mathbf{u}_{T+1})}{c_T(\mathbf{u}_T)} dF_{T+1}(y_{T+1}) \tag{4.11}$$

*where $c_T(\mathbf{u}_T)$ and $c_{T+1}(\mathbf{u}_{T+1})$ are respectively defined in (4.9) and (4.10), and that $F_{T+1}$ is a known marginal distribution function of $Y_{T+1}$.*

*Proof.* It is clear from the definition of conditional density, that if it exists, we must have

$$f_{Y_{T+1}|Y_\mathbf{T}}(y_{T+1}|\mathbf{y}_T) = \frac{h_{T+1}(\mathbf{y}_{T+1})}{h_T(\mathbf{y}_T)}.$$

The numerator can be written as

$$h_{T+1}(\mathbf{y}_{T+1}) = \frac{\partial^{T+1} C_{T+1}(\mathbf{u}_{T+1})}{\partial u_1 \ldots \partial u_T \partial u_{T+1}} \times \prod_{t=1}^{T+1} f_t(y_t) = c_{T+1}(\mathbf{u}_{T+1}) \times \prod_{t=1}^{T+1} f_t(y_t),$$

where $\mathbf{u}_{T+1}$ is understood to be evaluated at the respective marginals $F_t(y_t)$ for $t = 1, \ldots, T, T+1$. Similarly, we have

$$h_T(\mathbf{y}_T) = c_T(\mathbf{u}_T) \times \prod_{t=1}^{T} f_t(y_t).$$

From these, we now have

$$\begin{aligned}\mathbb{E}\left[Y_{T+1}|\mathbf{Y}_T = \mathbf{y}_T\right] &= \int y_{T+1} f_{Y_{T+1}|Y_\mathbf{T}}(y_{T+1}|\mathbf{y}_T) dy_{T+1} = \int y_{T+1} \frac{h_{T+1}(\mathbf{y}_{T+1})}{h_T(\mathbf{y}_T)} dy_{T+1} \\ &= \int y_{T+1} \frac{c_{T+1}(\mathbf{u}_{T+1}) \times \prod_{t=1}^{T+1} f_t(y_t)}{c_T(\mathbf{u}_T) \times \prod_{t=1}^{T} f_t(y_t)} dy_{T+1}\end{aligned}$$

and the result given in (4.11) follows. $\square$

Indeed, if we define $f^Q_{T+1}(y_{T+1}) = \dfrac{c_{T+1}(\mathbf{u}_{T+1})}{c_T(\mathbf{u}_T)} f_{T+1}(y_{T+1})$, we can verify that $dF^Q_{T+1} = f^Q_{T+1}(y_{T+1}) \cdot dy_{T+1}$ becomes a probability measure because both $c_T(\mathbf{u}_T) > 0$ and $f_{T+1}(y_{T+1}) > 0$ for any value of $T$. More specifically, we have

$$\begin{aligned}
\int dF^Q_{T+1} &= \int f^Q_{T+1}(y_{T+1}) dy_{T+1} = \int \frac{c_{T+1}(\mathbf{u}_{T+1})}{c_T(\mathbf{u}_T)} f_{T+1}(y_{T+1}) dy_{T+1} \\
&= \int \frac{h_{T+1}(\mathbf{y}_{T+1}) \times \prod_{t=1}^{T} f_t(y_t)}{h_T(\mathbf{y}_T) \times \prod_{t=1}^{T+1} f_t(y_t)} f_{T+1}(y_{T+1}) dy_{T+1} = \int f_{Y_{T+1}|\mathbf{Y_T}}(y_{T+1}|\mathbf{y}_T) dy_{T+1} = 1
\end{aligned}$$

In general, we can derive the copula structure based on a random effect framework as follows:

$$C_T(u_1, \dots, u_T) = \int \prod_{t=1}^{T} F_{Y_t|\theta}\left(F_t^{-1}(u_t)\right) \pi(\theta) d\theta = \mathbb{E}_\theta \left[\prod_{t=1}^{T} F_{Y_t|\theta}\left(F_t^{-1}(u_t)\right)\right],$$

where $F_{Y_t|\theta}$ and $F_t$ denotes the conditional and marginal distribution functions of $Y_t$, respectively.

## 4.4 GB2 and Bayesian credibility premium

Note that when $Y_t \sim \mathrm{GB2}(k, c, \psi, p)$, we derive the marginal distribution $F_t$ as follows. By letting $z = y^p$ and $v = \dfrac{z}{z + c^p}$, we have

$$\begin{aligned}
F_t(y_t) &= \int_0^{y_t} \frac{p}{yB(\psi, k)} \frac{y^{p\psi} c^{pk}}{[y^p + c^p]^{\psi+k}} dy = \int_0^{y_t^p} \frac{1}{zB(\psi, k)} \frac{z^{\psi} c^{pk}}{[z + c^p]^{\psi+k}} dz \\
&= \int_0^{\frac{y_t^p}{c^p + y_t^p}} \frac{1}{B(\psi, k)} v^{\psi-1}(1 - v)^{k-1} dv = B_{\psi,k}\left(\frac{y_t^p}{c^p + y_t^p}\right),
\end{aligned}$$

where $B_{\psi,k}(t)$ is the distribution function of Beta$(\psi, k)$ distribution so that $F_t^{-1}(u_t) = c\left(\dfrac{B^{-1}_{\psi,k}(u_t)}{1 - B^{-1}_{\psi,k}(u_t)}\right)^{1/p}$.

Therefore, the GB2 copula can be derived in the same manner as follows:

$$
\begin{aligned}
C_{k,\psi}(u_1,\ldots,u_T) &= \int \prod_{t=1}^{T} F_{Y_t|\theta}\left(c\left(\frac{B_{\psi,k}^{-1}(u_t)}{1-B_{\psi,k}^{-1}(u_t)}\right)^{1/p}\right)\pi(\theta)d\theta \\
&= \mathbb{E}_\theta\left[\prod_{t=1}^{T} F_{Y_t|\theta}\left(c\left(\frac{B_{\psi,k}^{-1}(u_t)}{1-B_{\psi,k}^{-1}(u_t)}\right)^{1/p}\right)\right],
\end{aligned} \tag{4.12}
$$

and its corresponding density is

$$
\begin{aligned}
c_{k,\psi}(u_1,\ldots,u_T) &= \frac{\partial^T}{\partial u_1\ldots\partial u_T}\int \prod_{t=1}^{T} F_{Y_t|\theta}\left(F_t^{-1}(u_t)\right)\pi(\theta)d\theta \\
&= \frac{\int \prod_{t=1}^{T} f_{Y_t|\theta}\left(F_t^{-1}(u_t)\right)\pi(\theta)d\theta}{\prod_{t=1}^{T} f_t\left(F_t^{-1}(u_t)\right)} = \frac{\int \prod_{t=1}^{T} f_{Y_t|\theta}(c\cdot q_t^{1/p})\pi(\theta)d\theta}{\prod_{t=1}^{T} f_t(c\cdot q_t^{1/p})} = \frac{h_T(c\cdot \mathbf{q}_T^{1/p})}{\prod_{t=1}^{T} f_t(c\cdot q_t^{1/p})} \\
&= \frac{B(\psi,k)^T}{B(\psi_T)}\frac{\prod_{t=1}^{T}(1+q_t)^{\psi+k}}{\left(1+\sum_{t=1}^{T} q_t\right)^{\psi T+k}} = \frac{\Gamma(k)^{T-1}\Gamma(\psi T+k)}{\Gamma(\psi+k)^T}\frac{\prod_{t=1}^{T}(1+q_t)^{\psi+k}}{\left(1+\sum_{t=1}^{T} q_t\right)^{\psi T+k}}
\end{aligned} \tag{4.13}
$$

where $q_t^{1/p} = F_t^{-1}(u_t)/c = \left(\frac{B_{\psi,k}^{-1}(u_t)}{1-B_{\psi,k}^{-1}(u_t)}\right)^{1/p}$.

Note that if we substitute $F_t(y_t)$ for $u_t$ where $F_t(y_t)$ is a marginal distribution function of GB2, it turns out that $c_{k,\psi}(u_1,\ldots,u_T)$ becomes $h_T(\mathbf{y}_T)$, a joint density of multivariate GB2 distribution.

By Fubini's theorem, the bivariate GB2 copula is given as follows:

$$
\begin{aligned}
C_{k,\psi}(u_1,u_2) &= \int_0^{c\cdot q_1^{1/p}}\int_0^{c\cdot q_2^{1/p}}\left(\int_0^\infty f_{Y_1|\theta}(y_1|\theta)f_{Y_2|\theta}(y_2|\theta)\pi(\theta)d\theta\right)dy_2dy_1 \\
&= \int_0^{c\cdot q_1^{1/p}}\int_0^{c\cdot q_2^{1/p}}\left(\frac{p^2}{B(\psi,\psi,k)y_1y_2}\frac{(y_1/c)^{p\psi}(y_2/c)^{p\psi}}{[(y_1/c)^p+(y_2/c)^p+1]^{2\psi+k}}\right)dy_2dy_1 \\
&= \int_0^{q_1}\int_0^{q_2}\left(\frac{1}{B(\psi,\psi,k)}\frac{v_1^{\psi-1}v_2^{\psi-1}}{[v_1+v_2+1]^{2\psi+k}}\right)dv_2dv_1
\end{aligned}
$$

where $v_i = (y_i/c)^p$ for $i = 1, 2$. From the above derivation, we see that the GB2 copula only depends on $\psi$ and $k$, but not on $c$ and $p$. Figure 4.1 provides a comparison of the contour plots of bivariate GB2 copulas using different set of parameters. Both parameters $\psi$ and $k$

describe the strength of the relationship as seen in this figure. For example, for a fixed $k$, a higher value of $\psi$ implies stronger dependence and for a fixed $\psi$, a higher value of $k$ implies weaker dependence.

Now, by applying (4.13), we can get the following result which is a crucial step to derive the Bayesian credibility premium under the multivariate GB2 distribution model.

**Lemma 4.4.** *Suppose* $(Y_1, \dots, Y_{T+1})'$ *follows a multivariate GB2 distribution as given in* (4.5) *and* $c_T$ *is the density of GB2 copula. Then the following relationship holds:*

$$f_{T+1}(y_{T+1}) \frac{c_{T+1}(\boldsymbol{u}_{T+1})}{c_T(\boldsymbol{u}_T)}\bigg|_{\boldsymbol{u}=F(\boldsymbol{y})} = \frac{p}{y_{T+1}\, B(\psi, k_T)} \times \frac{{c^*_{T,p}}^{pk_T} y_{T+1}^{p\psi}}{({c^*_{T,p}}^{p} + y_{T+1}^{p})^{\psi+k_T}}$$

*where* $c^*_{T,p} = \left(c^p + \sum_{t=1}^T y_t^p\right)^{1/p} = c\left(1 + \sum_{t=1}^T (y_t/c)^p\right)^{1/p}$, *and* $k_T = k + \psi T$.

*Proof.* See appendix for details of the proof. □

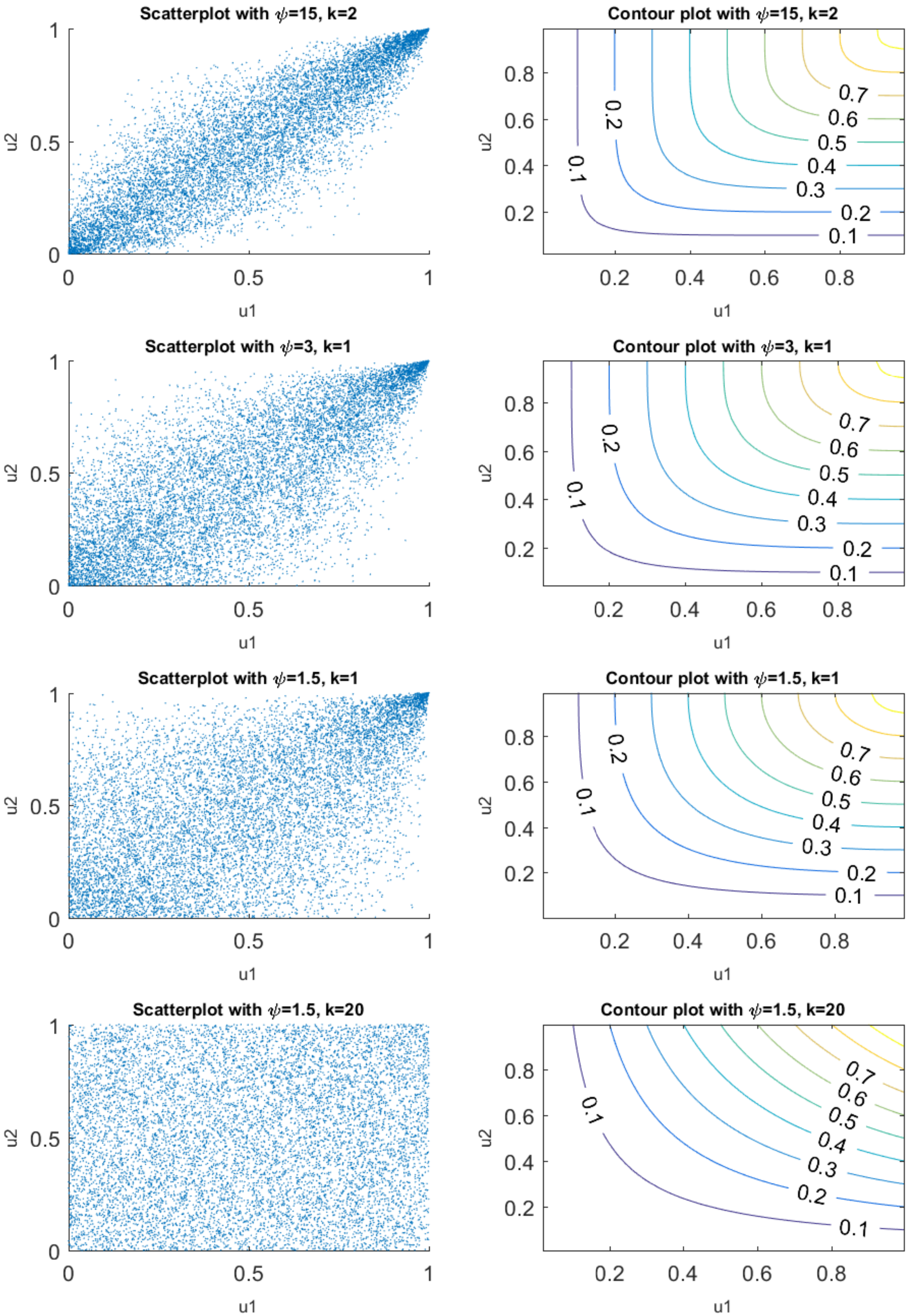
Based on Theorem 4.1, it is possible to evaluate the Bayesian credibility premium with GB2 copulas from the following theorem.

**Theorem 4.2.** *Suppose* $(Y_1, \dots, Y_{T+1})'$ *follows a multivariate distribution as described in (4.8) where* $C_T$ *is given as GB2 copula in (4.12) and the marginal distribution of* $Y_t$ *is given as* $F_t$*. Then the Bayesian credibility premium is written as follows:*

$$\mathbb{E}\left[Y_{T+1} \middle| \mathbf{Y}_T = \mathbf{y}_T\right] = \int_0^1 F_{T+1}^{-1}(u) \frac{(1+\eta_T)^{k_T}\left(\dfrac{1}{1-B_{\psi,k}^{-1}(u)}\right)^{\psi+k}}{\left(\eta_T + \dfrac{1}{1-B_{\psi,k}^{-1}(u)}\right)^{k_T+\psi}} \frac{B(k, k_T+\psi)}{B(k_T, k+\psi)} du, \qquad (4.14)$$

*where* $q_t = \left(\dfrac{B_{\psi,k}^{-1}(F_t(y_t))}{1-B_{\psi,k}^{-1}(F_t(y_t))}\right)$, $\eta_T = \sum_{t=1}^T q_t$, *and* $k_T = \psi T + k$.

Figure 4.1: Contour plots of bivariate GB2 copula with various parameterization



*Proof.* From Theorem 4.1 and (4.13), we can see that

$$\begin{aligned}
\mathbb{E}\left[Y_{T+1}|\mathbf{Y}_T=\mathbf{y}_T\right] &= \int y_{T+1}\frac{c_{T+1}(\mathbf{u}_{T+1})}{c_T(\mathbf{u}_T)}dF_{T+1}(y_{T+1}) \\
&= \int y_{T+1}\frac{\left(1+\sum_{t=1}^{T}q_t\right)^{k_T}(1+q_{T+1})^{\psi+k}}{\left(1+\sum_{t=1}^{T+1}q_t\right)^{k_T+\psi}}\frac{\Gamma(k)\Gamma(k_T+\psi)}{\Gamma(k_T)\Gamma(k+\psi)}dF_{T+1}(y_{T+1}) \\
&= \int y_{T+1}\frac{(1+\eta_T)^{k_T}(1+q_{T+1})^{\psi+k}}{(1+\eta_T+q_{T+1})^{k_T+\psi}}\frac{B(k,k_T+\psi)}{B(k_T,k+\psi)}dF_{T+1}(y_{T+1}).
\end{aligned} \tag{4.15}$$

Since $1+q_{T+1}=\left(\frac{1}{1-B^{-1}_{\psi,k}(F_{T+1}(y_{T+1}))}\right)$ and $F_{T+1}(y_{T+1})\sim\mathcal{U}[0,1]$, (4.15) can be expressed as follows:

$$\mathbb{E}\left[Y_{T+1}|\mathbf{Y}_T=\mathbf{y}_T\right]=\int_0^1 F^{-1}_{T+1}(u)\frac{(1+\eta_T)^{k_T}\left(\frac{1}{1-B^{-1}_{\psi,k}(u)}\right)^{\psi+k}}{\left(\eta_T+\frac{1}{1-B^{-1}_{\psi,k}(u)}\right)^{k_T+\psi}}\frac{B(k,k_T+\psi)}{B(k_T,k+\psi)}du.$$

□

As a special case of Theorem 4.2, it is possible to derive a nice closed form of Bayesian credibility premium when $\mathbf{Y}_T$ follows a multivariate GB2 distribution as follows.

**Corollary 4.1.** *Suppose $(Y_1,\dots,Y_{T+1})'$ follows a multivariate distribution as described in (4.8) where $C_T$ is given as GB2 copula in (4.12) and the marginal distribution of $Y_t$ is univariate GB2 distribution as in (4.6). Then the Bayesian credibility premium is given as follows:*

$$\mathbb{E}\left[Y_{T+1}|\mathbf{Y}_T=\mathbf{y}_T\right]=\left(1+\sum_{t=1}^{T}(y_t/c)^p\right)^{1/p}\frac{B(k,k_T-1/p)}{B(k-1/p,k_T)}\,\mu. \tag{4.16}$$

*Proof.* From Theorem 4.1 and Lemma 4.4, we can see that

$$f^Q_{T+1}(y_{T+1})=f_{T+1}(y_{T+1})\frac{c_{T+1}(\mathbf{u}_{T+1})}{c_T(\mathbf{u}_T)}\bigg|_{\mathbf{u}=F(\mathbf{y})}=\frac{p}{y_{T+1}\,B(\psi,k_T)}\times\frac{{c^*_{T,p}}^{pk_T}y^{p\psi}_{T+1}}{({c^*_{T,p}}^{p}+y^p_{T+1})^{\psi+k_T}}$$

This ratio of densities of the copula, $\dfrac{c_{T+1}(\mathbf{u}_{T+1})}{c_T(\mathbf{u}_T)}$, in fact induces a change of probability measure so that in effect, we can write the prediction as the following unconditional expectation

$$\mathbb{E}\left[Y_{T+1}|\mathbf{Y}_T = \mathbf{y}_T\right] = \mathbb{E}^Q\left[Y_{T+1}\right]$$

under a change of measure $dF^Q_{T+1}(y_{T+1}) = \dfrac{c_{T+1}(\mathbf{u}_{T+1})}{c_T(\mathbf{u}_T)}dF_{T+1}(y_{T+1})$. This change of measure allows us to construct an explicit expression for the posterior mean based on

$$\begin{aligned}
&Y_{T+1} \sim GB2(k, c, \psi, p) \qquad \text{under } dF_{T+1}(y_{T+1}) \\
&\ \text{and} \\
&Y_{T+1} \sim GB2(k_T, c^*_{T,p}, \psi, p) \quad \text{under } dF^Q_{T+1}(y_{T+1}).
\end{aligned}$$

Note that if $Y \sim GB2(k, c, \psi, p)$, then $\mathbb{E}\left[Y\right] = \dfrac{\Gamma(\psi + 1/p)}{\Gamma(\psi)}\dfrac{\Gamma(k - 1/p)}{\Gamma(k)}c$ from (4.7). Therefore, we can get the following result directly from the definition of $c^*_{T,p}$:

$$\begin{aligned}
\mathbb{E}\left[Y_{T+1}|\mathbf{Y}_T = \mathbf{y}_T\right] = \mathbb{E}^Q\left[Y_{T+1}\right] &= \frac{\Gamma(\psi + 1/p)}{\Gamma(\psi)}\frac{\Gamma(k_T - 1/p)}{\Gamma(k_T)}c^*_{T,p} \\
&= \left(1 + \sum_{t=1}^{T}(y_t/c)^p\right)^{1/p}\frac{\Gamma(\psi + 1/p)}{\Gamma(\psi)}\frac{\Gamma(k_T - 1/p)}{\Gamma(k_T)}c \\
&= \left(1 + \sum_{t=1}^{T}(y_t/c)^p\right)^{1/p}\ \frac{B(k_T - 1/p, k)}{B(k_T, k - 1/p)}\,\mu,
\end{aligned}$$

since $c = \dfrac{\Gamma(\psi)}{\Gamma(\psi + 1/p)}\dfrac{\Gamma(k)}{\Gamma(k - 1/p)}\,\mu$. □

If the marginal distribution of each $Y_t$ does now follow GB2, then the integral in (4.14) might not able to be evaluated analytically. The corollary below provides a nice approximation of the credibility premium based on Monte Carlo method.

**Corollary 4.2.** *A Monte Carlo approximation of Bayesian credibility premium in (4.14) is* *given as follows:*

$$\mathbb{E}\left[Y_{T+1}|\mathbf{Y}_T = \mathbf{y}_T\right] \simeq \frac{1}{M}\sum_{m=1}^{M} F_{T+1}^{-1}(u_{[m]}) \frac{(1+\eta_T)^{k_T}\left(\dfrac{1}{1-B_{\psi,k}^{-1}(u_{[m]})}\right)^{\psi+k}}{\left(\eta_T + \dfrac{1}{1-B_{\psi,k}^{-1}(u_{[m]})}\right)^{k_T+\psi}} \frac{B(k, k_T+\psi)}{B(k_T, k+\psi)}, \tag{4.17}$$

where $u_{[m]}$ for $m = 1, \ldots, M$ are random samples generated from $\mathcal{U}[0,1]$.

If $T = 0$, which means the case when there is no history of past claims, then $\eta_T = 0$ and $k_T = k$ so that (4.17) is reduced as follows:

$$\mathbb{E}\left[Y_{T+1}|\mathbf{Y}_T = \mathbf{y}_T\right] \simeq \frac{1}{M}\sum_{m=1}^{M} F_{T+1}^{-1}(u_{[m]}),$$

which is indeed a natural Monte Carlo approximation of the prior mean of $Y_{T+1}$, $\mathbb{E}\left[Y_{T+1}\right]$.

## 4.5 Numerical illustration - insurance ratemaking

For numerical illustration, we consider the case where our primary variable of interest, $y$, is the amount of claim for a portfolio of insurance contracts. In particular, we have a set of random claims for $T = 10$ periods: $y_1, \ldots, y_{10}$. Determining the pure premium based on historical claim experience is the subject of experience rating and credibility. In this case, the pure premium is the Bayesian credibility premium discussed in this paper. This numerical example shows how the choice of parameters affect Bayesian credibility premium with observed claim experience. To fix ideas, we control the prior mean, $\mu$, to be the same as 10 for all cases, but we vary the values of the corresponding coefficient of variation (CV), which is defined as $\sqrt{\mathbb{V}ar\left[Y\right]}/\mathbb{E}\left[Y\right]$. Table 4.1 shows different combinations of parameters which have the same mean but with different coefficient of variations, respectively.

Using these combinations of parameters, we assume three scenarios of observed claims, where each scenario is represented by the randomly generated quantiles of incurred claims. The

| CV | $\mu$ | $k$ | $\psi$ | c |
|---|---|---|---|---|
| 10% | 10 | 120.000 | 661.111 | 1.800 |
| 50% | 10 | 7.000 | 24.000 | 2.500 |
| 90% | 10 | 7.000 | 1.967 | 30.500 |
| 100% | 10 | 7.000 | 1.500 | 40.000 |
| 110% | 10 | 5.067 | 1.500 | 27.117 |
| 150% | 10 | 3.053 | 1.500 | 13.684 |
| 200% | 10 | 2.500 | 1.500 | 10.000 |

Table 4.1: Combinations of parameters for the GB2 posterior means

first is a 'risky' scenario so that the average of generated quantiles is 56.84%. The second is a 'normal' scenario so that the average of generated quantiles is 49.63%. Finally, the last is a 'safe' scenario so that the average of generated quantiles is 41.50%.

After assuming three scenarios based on the quantiles of the observed claims, we convert the quantile vectors to the observed claims under GB2 distribution for each set of parameters. We know that the 'weight factor' part of the GB2 credibility premium of the pure premium is given as $\left(1+\sum_{t=1}^{T}(y_t/c)^p\right)^{1/p} \frac{B(k_T-1/p,k)}{B(k_T,k-1/p)}$ from Corollary 4.1. Therefore, the value of weight factor is determined both by the parameters and the observed claim values. Note that even though we have the same quantiles of generated claims, the generated claim amount can still vary along with the assumed set of parameters.

Table 4.2 shows the result of the weight factors under all scenarios and parameter assumptions. Here, $w_H$ refers to the weight factors under the 'risky' scenario, $w_M$ to the weight factors under the 'normal' scenario, and $w_L$ to the weight factors under the 'safe' scenario, respectively.

This numerical illustration presents some very intuitively interesting results. From Table 4.2, we can infer that as we have higher coefficient of variation, the impact of credibility weighing factor increases. Moreover, for a policyholder with relatively higher claim experience, the resulting credibility weighing factor is greater than 1, which implies a penalty to policyholders with unfavorable claim experience. On the other hand, for a policyholder with relatively

| CV | $w_H$ | $w_M$ | $w_L$ |
|---|---|---|---|
| 10% | 1.0353 | 0.9972 | 0.9685 |
| 50% | 1.1996 | 0.9997 | 0.8508 |
| 90% | 1.2828 | 1.0051 | 0.7963 |
| 100% | 1.2933 | 1.0069 | 0.7919 |
| 110% | 1.3633 | 1.0109 | 0.7551 |
| 150% | 1.5340 | 1.0195 | 0.6767 |
| 200% | 1.6358 | 1.0211 | 0.6319 |

Table 4.2: Weight factors of the GB2 posterior mean under all scenarios and parameter assumptions

favorable claim experience, the credibility weighing factor is less than 1, which implies a bonus to the policyholder with more favorable claim experience.

## 4.6 Special Case: Generalized Pareto

Generalized Pareto (GP) distribution is a special case of GB2 distribution when $p = 1$. We may derive the multivariate GP distribution based on the Bayesian specification in Section 4.2 with $p = 1$ as follows:

$$Y_t|\Theta \overset{i.i.d.}{\sim} \text{Gamma}\left(\psi, \frac{\Theta}{\psi}\right) \quad \text{and} \quad \Theta \sim \text{I-Gamma}(k, \mu(k-1))$$

where

- Gamma $\left(\psi, \frac{\Theta}{\psi}\right)$ refers to a gamma with density

$$f_{Y_t}(y|\theta) = \frac{1}{y\Gamma(\psi)} \left(\frac{y\psi}{\theta}\right)^{\psi} \exp\left[-\left(\frac{y\psi}{\theta}\right)\right]$$

  and

- I-Gamma$(k, \mu(k-1))$ refers to an inverse-gamma with density

$$p(\theta) = \frac{1}{\theta\Gamma(k)} \left(\frac{\mu(k-1)}{\theta}\right)^k \exp\left[-\left(\frac{\mu(k-1)}{\theta}\right)\right].$$

By integrating out $\Theta$, we obtain the following multivariate GP distribution:

$$h_T(\mathbf{y}_T) = \frac{1}{B(\Psi_T)\prod_{t=1}^{T} y_t} \frac{c^k \prod_{t=1}^{T} y_t^{\psi}}{\left[\sum_{t=1}^{T} y_t + c\right]^{\psi T + k}},$$

where $c = \mu(k-1)/\psi$. Since $Y_t|\Theta$ are i.i.d., the marginal distribution is straightforward to determine by setting $T = 1$ so that

$$f_t(y_t) = \frac{1}{y_t\, B(\psi, k)} \frac{c^k\, y_t^{\psi}}{[y_t + c]^{\psi + k}}, \tag{4.18}$$

For a random variable $Y_t$ with density given in equation (4.18), we can write $Y_t \sim \text{GP}(k, c, \psi)$.

The following unconditional moments are straightforward to derive:

Mean: $\mathbb{E}[Y_t] = \mu$

Variance: $\text{V}ar\,[Y_t] = \mu^2 \left[\frac{k-1}{k-2}\frac{\psi+1}{\psi} - 1\right]$

Covariance: For $t \neq t'$, $\text{Cov}(Y_t, Y_{t'}) = \dfrac{\mu^2}{k-2}$

By Lemma 2, we see that if $(Y_1, \ldots, Y_T)'$ follows a multivariate GP distribution, then

$$\text{Corr}(Y_t, Y_{t'}) = \begin{cases} 1, & \text{if } \ t = t' \\ \rho_{1,k,\psi}, & \text{if } \ t \neq t' \end{cases} \qquad \text{where } \ \rho_{1,k,\psi} = \frac{\frac{k-1}{k-2} - 1}{\frac{k-1}{k-2}\frac{\psi+1}{\psi} - 1}.$$

Again, we can check that $\lim_{k\to\infty} \text{Cov}(Y_t, Y_{t'}) = \lim_{k\to\infty} \rho_{1,k,\psi} = 0$ if $t \neq t'$. The parameter $k$ gives a measure of the degree of correlation between pairs of GP random variables. Larger values of $k$ imply uncorrelated variables.

Under the GP model specification, from (4.14), the Bayesian posterior mean is given as

$$\mathbb{E}\left[Y_{T+1}|\mathbf{Y}_T = \mathbf{y}_T\right] = \left(1+\sum_{t=1}^{T}(y_t/c)\right)\frac{k-1}{k_T-1}\mu = \frac{(k-1)\mu+\psi T\,\overline{Y}}{\psi T+k-1} = \frac{\psi T}{\psi T+k-1}\cdot\overline{Y}+\frac{k-1}{\psi T+k-1}\cdot\mu,$$

which can be directly derived from the GB2 framework by letting $p = 1$. Interestingly, this has the form of a weighted average of prior mean $\mu$ and the average of previous observations, as shown in Bailey (1950), Mayerson (1964), and Bühlmann (1967). Note that this is not at all surprising and is a natural result because it is well known that if $Y_t|\Theta$ follows a distribution that belongs to the exponential family (Gamma distribution clearly belongs to the exponential family), then the posterior mean is exactly a linear combination of prior mean and sample mean of previous observations. For details of such result, see Jewell (1974).

It is well known that gamma distribution is a choice for modeling the severity component of property and casualty insurance claims. The usual model specification entertained by many insurers is as follows:

$$Y_t \overset{i.i.d.}{\sim} \text{Gamma}(\psi, \mu/\psi),$$

so that $\mathbb{E}\left[Y_t\right] = \mu$ and $\mathrm{V}ar\left[Y_t\right] = \mu^2/\psi$. We may regard this formulation as a limiting case of multivariate GP distribution because as $k \to \infty$,

$$\text{Cov}(Y_t, Y_{t'}) \to 0 \;\text{ and }\; \Theta \to 1.$$

An insurance company may wish to use the multivariate GP distribution as a predictive claims model by carefully calibrating the value of $k$.

## 4.7 Conclusion

As stated in Morris (1983), the concept of Bayesian shrinkage estimation is not limited to a normally distributed random variable or a member of the exponential family. This article extends the literature by developing explicit forms of Bayesian credibility premium within the family of GB2 copulas. The development is based on a new concept of using change of probability measure for copulas; this result is stated in Theorem 4.1. This theorem is the fundamental foundation for developing the explicit forms. Such credibility premium are very useful in actuarial science and insurance for experience-based ratemaking, where contractholders may be rewarded or penalized depending on their own claim experience. The concepts in this paper can be readily applied and extended in several ways. First, note that it is possible to use regression function $g(\mathbf{x}\beta)$ as prior mean, instead of the grand mean $\mu$. Second, credibility premium with GB2 copulas, but with non-GB2 marginals, can be obtained so long as the random variables have continuous support as shown in Theorem 4.2 and Corollary 4.2. Finally, as in many diverse applications, the proposed credibility premium can have a wide ranging applications for various multiparameter inference and regression problems.

# Chapter 5

# Multi-year microlevel dependent collective risk model via shared random effect

## 5.1 Introduction

According to Klugman et al. (2012), the aggregate loss in the classical collective risk model is defined as $S = \sum_{i=1}^{N} Y_i$, where $N$ means the number of claim and $Y_i$ denotes $i^{th}$ individual claim amounts over a fixed period of time with the following assumptions:

1. Conditional on $N = n$, the random variables $Y_1, Y_2, \ldots, Y_n$ are i.i.d. random variables.
2. Conditional on $N = n$, the common distribution of the random variables $Y_1, Y_2, \ldots, Y_n$ does not depend on $n$.
3. The distribution of $N$ does not depend in any way on the values of $Y_i$.

These assumptions might be convenient in terms of computational ease, however, such simplifying assumptions often lead to bias issues especially when used for risk classification.

In relaxing such assumptions, various models have been proposed in the insurance literature. An interesting method to model the dependence in the collective risk model is the so-called *two-part dependent frequency-severity model* as suggested by Frees et al. (2014a). In this model, the dependence is incorporated by using frequency as an explanatory variable in the severity component. A similar approach has been used by Frees et al. (2011a) in the modeling and prediction of frequency and severity of health care expenditure. Shi et al. (2015) suggested a three-part framework in order to capture the association between frequency and severity components. When generalized linear models (GLMs) are used with the number of claims treated as a covariate in claims severity, Garrido et al. (2016) showed that the pure premium includes a correction term for inducing dependence. When analyzing bonus-malus data, an interesting observation was made by Park et al. (2018) that dependence between claim frequency and severity is driven by the desire to reach a better bonus-malus class.

Applications of copula methods to capture dependence have been recently used in collective risk models. A majority of work in this area focused on modeling the dependence between frequency and average severity with parametric copulas. For example, Czado et al. (2012) used Gaussian copulas to extend traditional compound Poisson-Gamma two-part model and incorporated possible dependence. Krämer et al. (2013) suggested a similar joint copula-based approach and interestingly observed that ignoring dependence causes a severe underestimation of total loss in a portfolio. Frees et al. (2016b) extended the copula-based approach to dependent frequency and average severity using claims data with multiple lines of insurance business. While their findings suggested weak association between frequency and average severity, they concluded that there are strong dependencies among the lines of business.

Unlike choosing a suitable family of marginal distributions, it is usually much harder to choose the correct family of copulas when calibrating these dependent models with data. The work of Krämer et al. (2013) investigated test procedures for the selection of a suitable family of copulas in a dependent frequency and average severity model. However, Oh et al. (2019) illustrated that indeed it is even more difficult to choose the appropriate dependence structure between frequency and average severity that includes the classical collective risk model as a special case. In particular, even under the most naive assumption of independence between frequency and individual severities, choosing the correct parametric copula presents some challenges. Inspired by this phenomenon, Oh et al. (2019) and Cossette et al. (2019) discussed the construction of single year collective risk models with microlevel data to provide a suitable dependence structure between the frequency and severity components. In part, the extension in this paper that captures dependence of various types of dependence between claim frequency and claim severity over multiple years is motivated by the work of Oh et al. (2019).

In insurance industry, it is important to model the longitudinal property of the insurance losses to predict the fair premium in the future based on each policyholder's historical claims information. However, the existing copula methods in the literature cannot be directly applied in prediction of the premium due to at least one of the following difficulties:

- Limited to the analysis of data over a single period or cross-sectional data,
- The choice of the copula family to provide a suitable dependence structure between claim frequency and average claim severity can be difficult.

Alternatively, the random effect model can be used to model the longitudinal property of the insurance losses. Hernández-Bastida et al. (2009) and Oh et al. (2020) used the shared random effects model to construct the dependence in a collective risk model, where independence between claim frequency and severity conditional on the random effect is assumed and the dependence structure is naturally derived by the shared random effects. Jeong and Valdez (2020b) derived a closed form of credibility premium for compound loss which captures not only the dependence between frequency and severity but also dependence among the multi-year claims of the same policyholder. However, it is known that the overdispersion and serial dependence can be compounded in the random effect model. Such compounded effect of the random effect can possibly result in pseudo or fake dependence structure in the claims, which in turn leads to the poor prediction of the premium (Denuit et al., 2007; Murray et al., 2013; Lee et al., 2020).

In this regard, as a natural extension of shared random effects model and one-year dependent compound risk model, we propose a multi-year framework with microlevel data so that we may incorporate the following dependencies simultaneously:

- dependence between a frequency and a severity within a year,
- dependence between two distinct severities within a year,
- dependence among frequencies across years,
- dependence between a frequency and a severity in different years,
- dependence between two severities in different years.

Specifically, we use a factor copula representation, which can be viewed as a copula model version of the random effect model (Krupskii and Joe, 2013, 2015), by using 1-year microlevel model as building blocks.

The remainder of this paper is organized as follows. In Section 5.2, we propose a generalized shared random effects framework for multi-year microlevel collective risk model that incorporates all types of dependencies previously described. We demonstrate that previous methods for dependence modeling can be considered as special cases of our proposed model. In Section 5.3, we provide a concrete example of our proposed model with elliptical copulas. Because of simplicity, we focus on the family of Gaussian copulas to further explore various correlation structures that satisfy our framework. The performance of our proposed method is shown with a simulation study in Section 5.4. In Section 5.5, an empirical analysis with a special case of our proposed model is conducted with a dataset from an automobile insurance company. Concluding remarks are provided in Section 5.6 with some future directions of research.

## 5.2 Construction of the shared random effect parameter model

### 5.2.1 A motivating illustration

While copula methods are flexible in modeling the dependence, the “actual" flexibility comes from the proper choice of the parametric copula family. Although one may consider using the nonparametric copula method for the full flexibility in choosing a copula structure, modeling and interpreting dependence based on the non-parametric copula can be difficult as long as the discrete random variables are involved mainly due to the lack of uniquness (Genest and Nešlehová, 2007). While recent study in Yang et al. (2019) provides the safe copula estimation method for discrete outcomes in a regression context, it is known to suffer from the so-called curse of dimensionality.

Indeed, as shown in Oh et al. (2019), it is difficult to choose a proper parametric copula family for the frequency and average severity even under the most naive assumption, the case where frequency and individual severities are independent. This subsection summarizes the example in Oh et al. (2019) to explain such difficult and the necessity to use microlevel claims information.

Consider the classical collective risk model where frequency $N$ and the individual severity $Y_j$s are assumed to be independent. Further, assume that $N$ is a positive integer valued random variable with

$$\mathbb{P}\left(N=n\right)=\frac{1}{5}, \quad \text{for} \quad n=1,2,3,4,5,$$

and

$$Y_1,\cdots,Y_N \Big| N \overset{\text{i.i.d.}}{\sim} \text{Gamma}(\xi,\psi). \tag{5.1}$$

Then, (5.1) implies

$$M \Big| N \sim \text{Gamma}(\xi, \psi/N).$$

Clearly, $N$ and $M$ are not independent even though frequency and individual severities are independent. Since $N$ is discrete, the visualization and interpretation of the corresponding copula density function for $(N, M)$ can be difficult. Alternatively, Oh et al. (2019) provides the density function for the jittered version of $(N, M)$ as shown in Figure 5.1 where x-axis and y-axis corresponds to frequency $N$ and the average severity $M$, respectively.

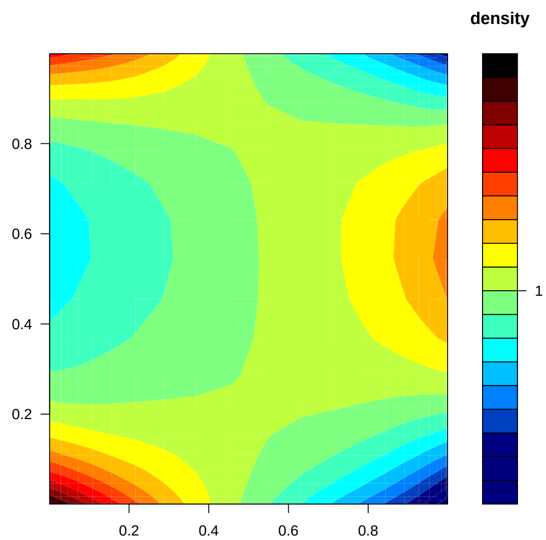



Figure 5.1: Contour plot, in Oh et al. (2019), of jittered copula density corresponding to $(N, M)$ using a kernel density estimation

Let $(U_1, U_2)$ be a bivariate random vector sampled from the copula of the jittered version of $(N, M)$. As shown in Figure 5.1, the density of the copula tends to be smaller in the middle part of $U_2$ when $U_1$ is smaller, wheras the density tends to be larger in the middle part of $U_2$ when $U_1$ is larger. Therefore, it is straightforward to see that conditional variance of $M$ decreases as $N$ increases in Figure 5.1, which is quite intuitive since $\mathbb{V}ar\left[M|N\right] = \xi^2\psi/N$ in this case.

This example illustrates that we can see that most existing copulas, including Gaussian and Archimedean copulas, are unable to accommodate the dependence between frequency and average severity properly. This is a motivation for the modeling the dependence based on the microlevel claims information rather than summarized claims information. We refer the readers to Oh et al. (2019) for more details of this example and the detailed construction of the jittered version of $(N, M)$.

### 5.2.2 Data structure and model specification

For non-life insurance, claims observed are typically a history of frequencies and severities for multiple years. For a policyholder observed for $\tau$ years, we have $n_1, \cdots, n_\tau$ which stand for frequency for each year, and corresponding individual severities $(\boldsymbol{y}_1, \cdots, \boldsymbol{y}_\tau)$ where

$$\boldsymbol{y}_t = \begin{cases} \text{not defined,} & n_t = 0; \\ (y_{t,1}, \cdots, y_{t,n_t}), & n_t > 0; \end{cases}$$

We find it convenient to define the following symbols for the description of data.

Define a random vector of length $N_t + 1$

$$\boldsymbol{Z}_t := \begin{cases} (N_t, Y_{t,1}, \cdots, Y_{t,N_t}), & N_t \geq 1; \\ 0, & N_t = 0, \end{cases}$$

and the realization of $\boldsymbol{Z}_t$ is denoted as

$$\boldsymbol{z}_t := \begin{cases} (n_t, \boldsymbol{y}_t), & n_t \geq 1; \\ 0, & n_t = 0. \end{cases}$$

Furthermore, multi-year extension of $\boldsymbol{Z}_t$ is defined as

$$\boldsymbol{Z}_{(\tau)} := (\boldsymbol{Z}_1, \cdots, \boldsymbol{Z}_\tau)$$

and the realization of $\boldsymbol{Z}_{(\tau)}$ is denoted as

$$\boldsymbol{z}_{(\tau)} := (\boldsymbol{z}_1, \cdots, \boldsymbol{z}_\tau).$$

In the subsequent, we describe a shared random effect parameter model for modeling the type of claims data we observe that primarily consist of frequencies and severities for multiple years.

**Model 5.1** (The copula linked shared random effect model)**.** *Consider the following random effect model for $\boldsymbol{Z}_t$ where the joint distribution between the observed losses and the shared random effect is presented with copulas.*

*i. Shared random effect $R$ follows a probability distribution with density $\pi$.*

*ii. Conditional on $R = r$, we have that $\boldsymbol{Z}_t$ for $t = 1, \cdots$ are independent observations whose distribution function is given by*

$$H_t(\boldsymbol{z}_t|r) := C_{(\theta_3,\theta_4)}\left(F_t(n_t|r), G_{t,1}(y_{t,1}|r), \ldots, G_{t,n_t}(y_{t,n_t}|r)\right) \tag{5.2}$$

*where $F_t$ and $G_{t,j}$ means marginal cumulative distribution functions of $N_t$ and $Y_{t,j}$, respectively and $g_{t,j}$ means joint density function of $Y_{t,j}$. As a result, we have the following distribution function of $\boldsymbol{Z}_{(\tau)}$*

$$H(\boldsymbol{z}_{(\tau)}) := \int \prod_{t=1}^{\tau} H_t(\boldsymbol{z}_t|r)\, \pi(r)\mathrm{d}r.$$

*iii. The parameters $\theta_3$ and $\theta_4$ of the copula $C_{(\theta_3,\theta_4)}$ controls the independence between the frequency and severities and independence among individual severities, respectively, within a year so that we have*

$$h_t(\boldsymbol{z}_t|r) = f_t(n_t|r) g_t^{[\text{joint}]}(\boldsymbol{y}_t|r) \quad \textit{if and only if} \quad \theta_3 = 0,$$

*where*

$$g_t(\boldsymbol{y}_t|r) = \prod_{j=1}^{N_t} g_{t,j}(y_{t,j}|r) \quad \textit{if and only if} \quad \theta_4 = 0,$$

*and* $g_t^{[\text{joint}]}$ *means joint density function of* $\boldsymbol{Y}_t$*.*

*iv.* $N_t \perp R$ *for* $t = 1, \ldots$ *if and only if* $\theta_1 = 0$*.*

*v.* $\boldsymbol{y}_t \perp R$ *for* $t = 1, \ldots$ *if and only if* $\theta_2 = 0$*.*

Figure 5.2 illustrates the dependence structure of our proposed model. In this figure we show that shared random effect $R$ induces the types of dependence that are of interest to us. To illustrate, $R$ is linked to the number of claims across years, $(N_1, \ldots, N_\tau)$, through $C_{\theta_1}$ so that $\theta_1$ is a parameter which captures dependence among claim counts between years. Likewise, $R$ is linked to the individual amounts of claims across years, $(\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_\tau)$, through $C_{\theta_2}$ so that $\theta_2$ is a parameter which captures dependence among claim amounts within and across the years. Furthermore, $C_{\theta_1}$ combined with $C_{\theta_2}$ introduces the dependence between the claim counts and individual severities within and across the years.

While, via the shared random effect $R$, the parameters $\theta_1$ and $\theta_2$ universally capture dependence among the claims across the years, the other parameters $\theta_3$ and $\theta_4$ specifically capture dependence within the claims of the same year. That is, $\theta_3$ is a parameter which incorporates the dependence between the claim count and claim amounts within a year whereas $\theta_4$ is a parameter which incorporates the dependence among claim amounts within a year. Similarly, $\theta_3$ combined with $\theta_4$ affects the dependence between the claim counts and individual severities within the year. As a result, while dependence among the claims in different years are modeled by $(\theta_1, \theta_2)$ only, the dependence among the claims in the same year are modeled by both $(\theta_1, \theta_2)$ and $(\theta_3, \theta_4)$. Note that our framework is distinguished from some existing work on dependence modeling with copulas such as Shi and Yang (2018) and Lee and Shi (2019), where average severity in the form of summarized data was used for modeling and implicitly precluded independence among the individual severities within the same year.

The idea of our multi-year microlevel collective risk model is that the observed claim for year $t$, $Z_t$, are independent for $t = 1, \dots, \tau$ given the shared random effect $R = r$ described as follows:

$$h(\boldsymbol{z}_{(\tau)}|r) = \prod_{t=1}^{\tau} h_t(z_t|r) \xRightarrow{\theta_3=0} \prod_{t=1}^{\tau} \left[ f_t(n_t|r) g_t^{[\text{joint}]} (\boldsymbol{y}_t|r) \right] \xRightarrow{\theta_3=\theta_4=0} \prod_{t=1}^{\tau} \left[ f_t(n_t|r) \left\{ \prod_{j=1}^{n_t} g_{t,j} (y_{t,j}|r) \right\} \right],$$

and

$$\begin{aligned} h(\boldsymbol{z}_{(\tau)}) = \int h(\boldsymbol{z}_{(\tau)}|r)\pi(r)\mathrm{d}r &\xRightarrow{\theta_3=0} \int \prod_{t=1}^{\tau} \left[ f_t(n_t|r) g_t^{[\text{joint}]} (\boldsymbol{y}_t|r) \right] \pi(r)\mathrm{d}r \\ &\xRightarrow{\theta_3=\theta_4=0} \int \prod_{t=1}^{\tau} \left[ f_t(n_t|r) \left\{ \prod_{j=1}^{n_t} g_{t,j} (y_{t,j}|r) \right\} \right] \pi(r)\mathrm{d}r, \end{aligned} \tag{5.3}$$

which is straightforward from iii and iv of Model 5.1.

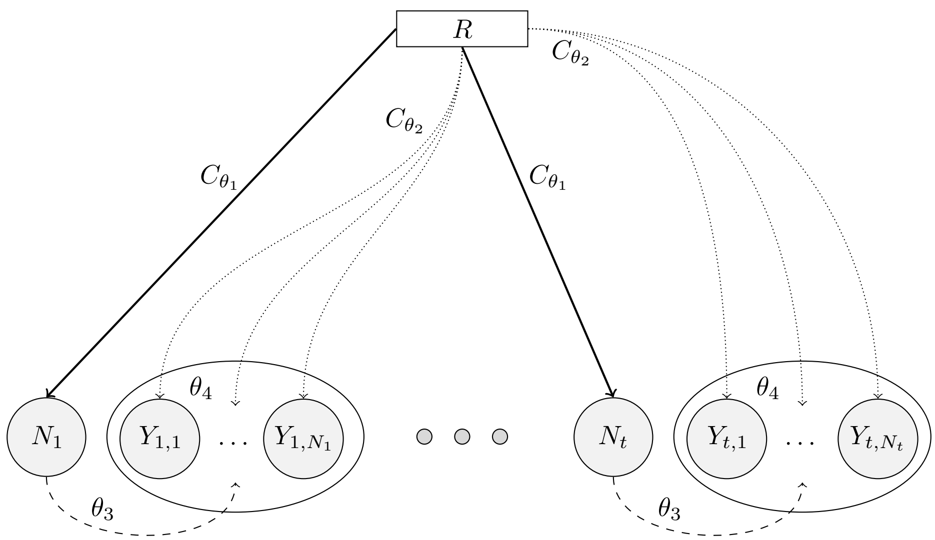



Figure 5.2: Visual representation of the multi-year microlevel shared random effect model

We note that this construction is similar to the model described by Krupskii and Joe (2013), which develops a factor copula model conditionally on a set of latent variables. In some sense, according to their paper, our approach leads to a one-factor copula model presented in Section 5.3. The primary difference in our approach is the clear intuitive interpretation of our model to describe the various types of dependence in a dependent collective risk model.

The well-definedness of Model 5.1 will also be discussed in Remark 5.1 in Section 5.3.

### 5.2.3 Special cases

It is immediate to see that the classical collective risk model of Klugman et al. (2012) is a special case of our proposed model where $\theta_1 = \theta_2 = \theta_3 = \theta_4 = 0$. This is the case when all frequencies and severities are mutually independent. Baumgartner et al. (2015) proposed shared random effects model to capture association between frequency and the average severity, which is just another special case of our proposed model. This is the case when $\theta_3 = \theta_4 = 0$. Finally, it is also easy to check that single-year microlevel collective risk model, proposed by Oh et al. (2019), is another special case of our proposed model. This is when $\theta_1 = \theta_2 = 0$. In this regard, our proposed framework is quite comprehensive that allows other dependence models that have appeared in the literature as special cases.

## 5.3 Factor copula model based on the elliptical distributions

Copulas generated by elliptical distributions, also called *elliptical copulas*, have the correlation matrix as the primary parameter describing dependence between the components. The Gaussian and $t$ copulas belong to the family of elliptical copulas. We refer to Landsman and Valdez (2003) for other choices of elliptical copulas including the copulas generated from multivariate Cauchy or multivariate logistic distribution. In this section, for simplicity, apparent ease of computations, and steering clear of distractions from the general case, we focus on the case of Gaussian copulas. In Appendix D, we illustrate how Gaussian copulas in multi-year microlevel collective risk model can be generalized into the elliptical copulas by providing an example of $t$ copula among other choices of elliptical copulas. Specifically, we consider Gaussian copulas with a specific covariance matrix to accommodate the dependence structure of multi-year microlevel collective risk model, and show that such Gaussian copula models can be represented as factor copula models. For the use in elliptical copulas including the Gaussian and $t$ copulas in mind, we begin with describing dependence structure via correlation matrices.

### 5.3.1 Dependence structure via correlation matrix

We start with definition of symbols. Denote $\mathbb{N}$, $\mathbb{N}_0$, $\mathbb{R}$, and $\mathbb{R}^+$ by the set of positive integer, the set of non-negative integer, the set of real number, and the set of positive real number, respectively.

For a $n \times m$ matrix $\boldsymbol{M}$, we denote $(i,j)$-th component of $\boldsymbol{M}$ as $[\boldsymbol{M}]_{ij}$. For a row vector $\boldsymbol{v}$ of length $n$, we denote the $i$-th component of $\boldsymbol{v}$ as $[\boldsymbol{v}]_i$. For $n \in \mathbb{N}$, define $\mathbf{1}_n$ and $\boldsymbol{J}_{n\times n}$ as a column vector of 1 with length $n$ and a $n \times n$ matrix of ones, respectively. We use $\boldsymbol{I}_n$ for $n \in \mathbb{N}$ to represent the $n \times n$ identity matrix.

Suppose $\mathbf{\Sigma}_{1,1}$, $\mathbf{\Sigma}_{1,2}$, $\mathbf{\Sigma}_{2,1}$, and $\mathbf{\Sigma}_{2,2}$ are $\ell \times \ell$, $\ell \times m$, $m \times \ell$, and $m \times m$ matrices, respectively. Define $(\ell + m) \times (\ell + m)$ matrix $\mathbf{\Sigma}$ as

$$\mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_{1,1} & \mathbf{\Sigma}_{1,2} \\ \mathbf{\Sigma}_{2,1} & \mathbf{\Sigma}_{2,2} \end{pmatrix}$$

If $\mathbf{\Sigma}_{2,2}$ is invertible, the Schur complement of the block $\mathbf{\Sigma}_{2,2}$ of the matrix $\mathbf{\Sigma}$ is the $\ell \times \ell$ matrix defined by

$$\mathbf{\Sigma} /\!\!/ \mathbf{\Sigma}_{1,1} := \mathbf{\Sigma}_{2,2} - \mathbf{\Sigma}_{2,1} \left(\mathbf{\Sigma}_{1,1}\right)^{-1} \mathbf{\Sigma}_{1,2}.$$

**Definition 5.1.** *For $\boldsymbol{n} = (n_1, \cdots, n_\tau) \in (\mathbb{N}_0)^\tau$ and $\boldsymbol{\rho} = (\rho_1, \cdots, \rho_5) \in \mathbb{R}^5$ define the partitioned matrix*

$$\mathbf{\Sigma}^{(\boldsymbol{\rho})}_{(\boldsymbol{n})} := \begin{pmatrix} \mathbf{\Sigma}^{(\boldsymbol{\rho})}_{11} & \cdots & \mathbf{\Sigma}^{(\boldsymbol{\rho})}_{1\tau} \\ \vdots & \ddots & \vdots \\ \mathbf{\Sigma}^{(\boldsymbol{\rho})}_{\tau 1} & \cdots & \mathbf{\Sigma}^{(\boldsymbol{\rho})}_{\tau\tau} \end{pmatrix}. \tag{5.4}$$

*For* $i = 1, \cdots, \tau$, *the matrix* $\mathbf{\Sigma}_{tt}^{(\boldsymbol{\rho})}$ *is a* $(n_t + 1) \times (n_t + 1)$ *matrix defined as*

$$\left[\mathbf{\Sigma}_{tt}^{(\boldsymbol{\rho})}\right]_{\ell m} = \begin{cases} 1, & \ell = m; \\ \rho_2, & \ell \neq m, \quad \min\{\ell, m\} \geq 2; \\ \rho_1, & \textit{elsewhere}; \end{cases}$$

*for* $\ell, m = 1, \cdots, n_t + 1$. *Furthermore, for* $i, j = 1, \cdots, t$ *with* $i \neq j$, *the matrix* $\mathbf{\Sigma}_{ij}$ *is a* $(n_i + 1) \times (n_j + 1)$ *matrix defined as*

$$\left[\mathbf{\Sigma}_{tj}^{(\boldsymbol{\rho})}\right]_{\ell m} = \begin{cases} \rho_3, & \ell = m = 1; \\ \rho_5, & \min\{\ell, m\} \geq 2; \\ \rho_4, & \textit{elsewhere}; \end{cases}$$

*for* $\ell = 1, \cdots, n_t + 1$ *and* $m = 1, \cdots, n_j + 1$.

**Example 5.1.** *Consider the case* $\boldsymbol{n} = (2, 3)$. *Then we can write out* $\mathbf{\Sigma}_{(\boldsymbol{n})}^{(\boldsymbol{\rho})}$ *by denoting* $\boldsymbol{n} = (2, 3)$ *and* $\boldsymbol{\rho} = (\rho_1, \cdots, \rho_5) \in \mathbb{R}^5$. *As a result,* $\mathbf{\Sigma}_{(\boldsymbol{n})}^{(\boldsymbol{\rho})}$ *is a* $7 \times 7$ *defined as*

$$\mathbf{\Sigma}_{(\boldsymbol{n})}^{(\boldsymbol{\rho})} := \begin{pmatrix} \mathbf{\Sigma}_{11}^{(\boldsymbol{\rho})} & \mathbf{\Sigma}_{12}^{(\boldsymbol{\rho})} \\ \mathbf{\Sigma}_{21}^{(\boldsymbol{\rho})} & \mathbf{\Sigma}_{22}^{(\boldsymbol{\rho})} \end{pmatrix}$$

*where*

$$\mathbf{\Sigma}_{11}^{(\boldsymbol{\rho})} = \begin{pmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_1 & \rho_2 & 1 \end{pmatrix} \quad \mathbf{\Sigma}_{12}^{(\boldsymbol{\rho})} = \begin{pmatrix} \rho_3 & \rho_4 & \rho_4 & \rho_4 \\ \rho_4 & \rho_5 & \rho_5 & \rho_5 \\ \rho_4 & \rho_5 & \rho_5 & \rho_5 \end{pmatrix} \quad \textit{and} \quad \mathbf{\Sigma}_{22}^{(\boldsymbol{\rho})} = \begin{pmatrix} 1 & \rho_1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 & \rho_2 \\ \rho_1 & \rho_2 & 1 & \rho_2 \\ \rho_1 & \rho_2 & \rho_2 & 1 \end{pmatrix}.$$

*Furthermore, from the above and the following*

$$\boldsymbol{\Sigma}_{21}^{(\boldsymbol{\rho})} = \left(\boldsymbol{\Sigma}_{12}^{(\boldsymbol{\rho})}\right)^{\mathrm{T}},$$

*we have*

$$\boldsymbol{\Sigma}_{(\boldsymbol{n})}^{(\boldsymbol{\rho})} = \left(\begin{array}{ccc|cccc} 1 & \rho_1 & \rho_1 & \rho_3 & \rho_4 & \rho_4 & \rho_4 \\ \rho_1 & 1 & \rho_2 & \rho_4 & \rho_5 & \rho_5 & \rho_5 \\ \rho_1 & \rho_2 & 1 & \rho_4 & \rho_5 & \rho_5 & \rho_5 \\ \hline \rho_3 & \rho_4 & \rho_4 & 1 & \rho_1 & \rho_1 & \rho_1 \\ \rho_4 & \rho_5 & \rho_5 & \rho_1 & 1 & \rho_2 & \rho_2 \\ \rho_4 & \rho_5 & \rho_5 & \rho_1 & \rho_2 & 1 & \rho_2 \\ \rho_4 & \rho_5 & \rho_5 & \rho_1 & \rho_2 & \rho_1 & 1 \end{array}\right)$$

In the matrix $\boldsymbol{\Sigma}_{(\boldsymbol{n})}^{(\boldsymbol{\rho})}$, each component will be used for modeling the correlation between frequencies and severities within and across years. For example, the partitioned matrix $\boldsymbol{\Sigma}_{tt}^{(\boldsymbol{\rho})}$ is a $(n_t+1)\times(n_t+1)$ matrix describing the correlation structure of the random vector $(N_t, Y_{t,1}, \cdots, Y_{t,n_t})$. Specifically, $\rho_1$ in $\boldsymbol{\Sigma}_{tt}^{(\boldsymbol{\rho})}$ is used for a correlation between a frequency and a severity in the $t$-th year, and $\rho_2$ in $\boldsymbol{\Sigma}_{tt}^{(\boldsymbol{\rho})}$ is used for a correlation among the severities in the $t$-th year. On the other hand, the partitioned matrix $\boldsymbol{\Sigma}_{tj}^{(\boldsymbol{\rho})}$ is a $(n_t+1)\times(n_j+1)$ matrix describing the correlation structure between the random vectors $(N_t, Y_{t,1}, \cdots, Y_{t,n_t})$ and $(N_j, Y_{j,1}, \cdots, Y_{j,n_j})$. Specifically, $\rho_3$ in $\left[\boldsymbol{\Sigma}_{tj}^{(\boldsymbol{\rho})}\right]_{11}$ is used for a correlation between the frequencies in the different years, and $\rho_4$ in $\boldsymbol{\Sigma}_{tj}^{(\boldsymbol{\rho})}$ is used for a correlation between a frequency in different years. Finally, $\rho_5$ in $\boldsymbol{\Sigma}_{tj}^{(\boldsymbol{\rho})}$ is used for a correlation between a frequency and a severity in different years. The following is summarization for the meaning of each correlation:

- $\rho_1$: correlation between a frequency and a severity within a year;
- $\rho_2$: correlation among two distinct severities within a year;
- $\rho_3$: correlation among frequencies across years;

- $\rho_4$: correlation between a frequency and a severity in different years;

- $\rho_5$: correlation between two severities in different years.

We finally note that $\mathbf{\Sigma}_{tt}^{(\boldsymbol{\rho})}$ only depends on $(\rho_1, \rho_2)$ while $\mathbf{\Sigma}_{tj}^{(\boldsymbol{\rho})}$ for $t \neq j$ only depends on $(\rho_3, \rho_4, \rho_5)$. Hence, we find that it is convenient to use $\mathbf{\Sigma}_{tt}^{(\boldsymbol{\rho}^*)}$ with $\boldsymbol{\rho}^* = (\rho_1, \rho_2)$ to stand for $\mathbf{\Sigma}_{tt}^{(\boldsymbol{\rho})}$, and similarly $\mathbf{\Sigma}_{tj}^{(\boldsymbol{\rho}^*)}$ for $t \neq j$ with $\boldsymbol{\rho}^* = (\rho_3, \rho_4, \rho_5)$ to stand for $\mathbf{\Sigma}_{tj}^{(\boldsymbol{\rho})}$ in a clear context.

**Definition 5.2.** *For $\boldsymbol{n} = (n_1, \cdots, n_\tau) \in \mathbb{N}_0^\tau$, $\boldsymbol{\rho} = (\rho_1, \cdots, \rho_5) \in (-1,1)^5$, $\boldsymbol{\theta} = (\theta_1, \theta_2) \in (-1,1)^2$, define the partitioned matrix as $\mathbf{\Sigma}_{(\boldsymbol{n})}^{(\boldsymbol{\rho};\boldsymbol{\theta})}$ as*

$$\mathbf{\Sigma}_{(\boldsymbol{n})}^{(\boldsymbol{\rho},\boldsymbol{\theta})} := \begin{pmatrix} \boldsymbol{I}_1 & \mathbf{\Omega}_{(\boldsymbol{n})}^{(\boldsymbol{\theta})} \\ \left(\mathbf{\Omega}_{(\boldsymbol{n})}^{(\boldsymbol{\theta})}\right)^{\mathrm{T}} & \mathbf{\Sigma}_{(\boldsymbol{n})}^{(\boldsymbol{\rho})} \end{pmatrix} \tag{5.5}$$

*where $\mathbf{\Sigma}_{(\boldsymbol{n})}^{(\boldsymbol{\rho})}$ is defined in (5.4) and $\mathbf{\Omega}_{(\boldsymbol{n})}^{(\boldsymbol{\theta})}$ is a $1 \times (\bar{\boldsymbol{n}} + \tau)$ matrix which can be expressed based on the following partitioned matrix*

$$\mathbf{\Omega}_{(\boldsymbol{n})}^{(\boldsymbol{\theta})} := \left(\mathbf{\Omega}_{n_1}^{(\boldsymbol{\theta})}, \cdots, \mathbf{\Omega}_{n_\tau}^{(\boldsymbol{\theta})}\right)$$

*with $\mathbf{\Omega}_{n_t}^{(\boldsymbol{\theta})}$ being a $1 \times (n_t + 1)$ matrix given by*

$$\left[\mathbf{\Omega}_{n_t}^{(\boldsymbol{\theta})}\right]_{1\ell} := \begin{cases} \theta_1, & \ell = 1; \\ \theta_2, & \textit{otherwise.} \end{cases}$$

In Definition 5.2, we have introduced two parameters $\theta_1$ and $\theta_2$. We impose natural dependence for multiples years of observed claims by using the shared random effect $R$, which will affect all frequency and severities in any calendar year. In this regard, $\theta_1$ will be served as correlation parameter between the random effect $R$ and a frequency, and $\theta_2$ will be served as correlation parameter between a random effect $R$ and each severity, as described in Figure 5.1.

**Example 5.2.** *Consider the case $\boldsymbol{n} = (2,3) \in \mathbb{N}_0^2$, then one can represent $\mathbf{\Sigma}_{\boldsymbol{n}}^{(\boldsymbol{\rho},\boldsymbol{\theta})}$ as a* *partitioned matrix as*

$$\boldsymbol{\Sigma}_{(\boldsymbol{n})}^{(\boldsymbol{\rho},\boldsymbol{\theta})} := \begin{pmatrix} \boldsymbol{I}_1 & \boldsymbol{\Omega}_{(\boldsymbol{n})}^{(\boldsymbol{\theta})} \\ \left(\boldsymbol{\Omega}_{(\boldsymbol{n})}^{(\boldsymbol{\theta})}\right)^{\mathrm{T}} & \boldsymbol{\Sigma}_{(\boldsymbol{n})}^{(\boldsymbol{\rho})} \end{pmatrix}$$

*where* $\boldsymbol{\Sigma}_{\boldsymbol{n}}^{(\boldsymbol{\rho})}$ *is in* (5.4)*, and*

$$\boldsymbol{\Omega}_{\boldsymbol{n}}^{(\boldsymbol{\theta})} = \begin{pmatrix} \theta_1 & \theta_2 & \theta_2 & \theta_1 & \theta_2 & \theta_2 & \theta_2 \end{pmatrix}.$$

*Hence, we have*

$$\boldsymbol{\Sigma}_{(\boldsymbol{n})}^{(\boldsymbol{\rho},\boldsymbol{\theta})} = \left(\begin{array}{c:ccc:cccc} 1 & \theta_1 & \theta_2 & \theta_2 & \theta_1 & \theta_2 & \theta_2 & \theta_2 \\ \hdashline \theta_1 & 1 & \rho_1 & \rho_1 & \rho_3 & \rho_4 & \rho_4 & \rho_4 \\ \theta_2 & \rho_1 & 1 & \rho_2 & \rho_4 & \rho_5 & \rho_5 & \rho_5 \\ \theta_2 & \rho_1 & \rho_2 & 1 & \rho_4 & \rho_5 & \rho_5 & \rho_5 \\ \hdashline \theta_1 & \rho_3 & \rho_4 & \rho_4 & 1 & \rho_1 & \rho_1 & \rho_1 \\ \theta_2 & \rho_4 & \rho_5 & \rho_5 & \rho_1 & 1 & \rho_2 & \rho_2 \\ \theta_2 & \rho_4 & \rho_5 & \rho_5 & \rho_1 & \rho_2 & 1 & \rho_2 \\ \theta_2 & \rho_4 & \rho_5 & \rho_5 & \rho_1 & \rho_2 & \rho_1 & 1 \end{array}\right).$$

Now, for $\boldsymbol{n} = (n_1, \cdots, n_\tau) \in (\mathbb{N}_0)^\tau$ and $\boldsymbol{\rho} = (\rho_1, \cdots, \rho_5) \in \mathbb{R}^5$, we consider reparameterization of a matrix $\boldsymbol{\Sigma}_{(\boldsymbol{n})}^{(\boldsymbol{\rho})}$ with

$$\begin{cases} \rho_1 = \theta_1\theta_2 + \theta_3\theta_4 \\ \rho_2 = \theta_2^2 + \theta_4^2 \\ \rho_3 = \theta_1^2 \\ \rho_4 = \theta_1\theta_2 \\ \rho_5 = \theta_2^2 \end{cases} \tag{5.6}$$

for

$$\boldsymbol{\theta} = (\theta_1, \cdots, \theta_4) \in \mathbb{R}^4.$$

The following theorem provides some results related with reparameterization in (5.6).

**Theorem 5.1.** *For* $\boldsymbol{n} = (n_1, \cdots, n_\tau) \in \mathbb{N}_0^\tau$, $\boldsymbol{\rho} = (\rho_1, \cdots, \rho_5) \in (-1,1)^5$, *consider the Schur Complement of the block* $\boldsymbol{I}_1$ *of the matrix* $\boldsymbol{\Sigma}_{(\boldsymbol{n})}^{(\boldsymbol{\rho},\boldsymbol{\theta})}$ *in* (5.5) *denoted as* $\boldsymbol{M} := \boldsymbol{\Sigma}_{(\boldsymbol{n})}^{(\boldsymbol{\rho},\boldsymbol{\theta})} /\!/ \boldsymbol{I}_1$. *For convenience, consider the following block matrix representation of* $\boldsymbol{M}$ *as*

$$\boldsymbol{M} = \begin{pmatrix} \boldsymbol{M}_{11} & \cdots & \boldsymbol{M}_{1\tau} \\ \vdots & \ddots & \vdots \\ \boldsymbol{M}_{\tau 1} & \cdots & \boldsymbol{M}_{\tau\tau} \end{pmatrix} \tag{5.7}$$

*where* $\boldsymbol{M}_{ij}$ *is a* $n_i \times n_j$ *matrix. Then, we have the following results.*

*i. For any* $\boldsymbol{n} \in (\mathbb{N}_0)^\tau$, $\boldsymbol{M}$ *is a block diagonal matrix, i.e.* $\boldsymbol{M}_{ij}$ *is a* $n_i \times n_j$ *zero matrix whenever* $i \neq j$, *if and only if* $\rho_3$, $\rho_4$, *and* $\rho_5$ *satisfy*

$$\rho_3 = \theta_1^2, \quad \rho_4 = \theta_1\theta_2, \quad \textit{and} \quad \rho_5 = \theta_2^2. \tag{5.8}$$

*ii. A matrix* $\boldsymbol{\Sigma}_{(\boldsymbol{n})}^{(\boldsymbol{\rho},\boldsymbol{\theta})}$ *is positive definite and* $\boldsymbol{M}$ *is a block diagonal matrix for any* $\boldsymbol{n} \in (\mathbb{N}_0)^\tau$ *if and only if* $\boldsymbol{\rho}$ *is represented as in* (5.6) *and satisfying*

$$\theta_1^2 + \theta_3^2 < 1 \quad \textit{and} \quad \theta_2^2 + \theta_4^2 < 1. \tag{5.9}$$

*iii. A matrix* $\boldsymbol{\Sigma}_{(\boldsymbol{n})}^{(\boldsymbol{\rho})}$ *with the parametrization in* (5.6) *is positive definite for any* $\boldsymbol{n} \in (\mathbb{N}_0)^\tau$ *if* $\boldsymbol{\theta}$ *satisfies* (5.9).

*Proof.* For the proof of part i, it suffices to show that if $i \neq j$, then

$$\boldsymbol{M}_{ij} = \boldsymbol{\Sigma}_{ij} - [\boldsymbol{\Omega}_{n_i}^{(\boldsymbol{\theta})}]^{\mathrm{T}} \boldsymbol{\Omega}_{n_j}^{(\boldsymbol{\theta})}$$

by definition of $\boldsymbol{M}$ where $\boldsymbol{\Sigma}_{ij}$ and $\boldsymbol{\Omega}_{n_t}^{(\boldsymbol{\theta})}$ are defined in (5.5) and (5.7), respectively and it can be written as follows:

$$[\boldsymbol{M}_{ij}]_{\ell m} = \begin{cases} \rho_3 - \theta_1^2, & \ell = m = 1; \\ \rho_5 - \theta_2^2, & \min\{\ell, m\} \geq 2; \\ \rho_4 - \theta_1\theta_2, & \text{elsewhere.} \end{cases}$$

For the proof of part ii, by Schur decomposition, we have $\boldsymbol{\Sigma}_{(\boldsymbol{n})}^{(\boldsymbol{\rho},\boldsymbol{\theta})}$ is positive definite if and only if $\boldsymbol{M}$ is positive definite. Since $\boldsymbol{M}$ is a block diagonal matrix provided (5.6) is satisfied, checking the positive definiteness of $\boldsymbol{M}$ is equivalent to check whether $\boldsymbol{M}_{jj}$ is positive definite or not. Hence, a matrix $\boldsymbol{\Sigma}_{(\boldsymbol{n})}^{(\boldsymbol{\rho},\boldsymbol{\theta})}$ is positive definite and $\boldsymbol{M}$ is a block diagonal matrix for any $\boldsymbol{n} \in (\mathbb{N}_0)^\tau$ if and only if $\boldsymbol{\Sigma}_{tt}^{(\boldsymbol{\rho}^*)}$ is positive definite for any $t \in \mathbb{N}_0$ where

$$\rho_1^* = \frac{\rho_1 - \theta_1\theta_2}{\sqrt{1-\theta_1^2}\sqrt{1-\theta_2^2}} \quad \text{and} \quad \rho_2^* = \frac{\rho_2 - \theta_2^2}{1-\theta_2^2}, \quad \theta_1, \theta_2 \in (-1, 1).$$

Following Corollary 1 in Oh et al. (2019), we have positive definite $\boldsymbol{\Sigma}_{tt}^{(\boldsymbol{\rho}^*)}$ for any $t \in \mathbb{N}_0$ if and only if

$$(\rho_1^*)^2 < \rho_2^* < 1. \tag{5.10}$$

Finally, simple argument shows that (5.10) with the condition $\theta_1, \theta_2 \in (-1, 1)$ is equivalent with

$$\rho_1 = \theta_1\theta_2 + \theta_3\theta_4 \quad \text{and} \quad \rho_2 = \theta_2^2 + \theta_4^2$$

for

$$\theta_1^2 + \theta_3^2 < 1 \quad \text{and} \quad \theta_2^2 + \theta_4^2 < 1.$$

The proof of part iii immediately follows from part i and ii. □

### 5.3.2 The special case of Gaussian copulas

Let $F_t$ be non-negative integer-valued distribution functions with the respective probability mass functions $f_t$ for $t \in \mathbb{N}$. Let $G_t$ and $G_{t,j}$ be non-negative real-valued distribution functions with respective probability densities $g_t$ and $g_{t,j}$ for $t, j \in \mathbb{N}$. While it is not necessary but for simplicity, we assume $G_{tj} = G_t$ for any $t, j \in \mathbb{N}$.

We use $\Phi$ and $\phi$ to denote the standard normal distribution and the corresponding density function, respectively. For a vector $\boldsymbol{\mu} \in \mathbb{R}^n$ and a $n \times n$ covariance matrix $\boldsymbol{\Sigma}$, we use $\Phi_{\boldsymbol{\mu},\boldsymbol{\Sigma}}$ to denote the distribution function of multivariate normal distribution with mean $\boldsymbol{\mu}$ and a covariance matrix $\boldsymbol{\Sigma}$, and $\phi_{\boldsymbol{\mu},\boldsymbol{\Sigma}}$ to denote the corresponding density function. Now, we are ready to present the multi-year microlevel collective risk model where the Gaussian copula is used to model the dependence.

**Model 5.2** (The Gaussian copula model for the multi-year microlevel collective risk model)**.** *Suppose $\boldsymbol{\rho}$ satisfies* (5.6)*. Then, consider the random vector $\boldsymbol{Z}_t$ whose joint distribution function $H(\boldsymbol{z}_\tau)$ is given by the following copula model representation*

$$H(\boldsymbol{z}_{(\tau)}) = C^{(\boldsymbol{\rho})}_{(\boldsymbol{n})}\left(F_1(n_1), G_{1,1}(y_{1,1}), \cdots, G_{1,n_1}(y_{1,n_1}), \cdots F_\tau(n_\tau), G_{\tau,1}(y_{\tau,1}), \cdots, G_{\tau,n_\tau}(y_{\tau,n_\tau})\right) \tag{5.11}$$

*where $C^{(\boldsymbol{\rho})}_{(\boldsymbol{n})}$ is a Gaussian copula with correlation matrix $\boldsymbol{\Sigma}^{(\boldsymbol{\rho})}_{(\boldsymbol{n})}$.*

From Lemma 5.1, the matrix $\boldsymbol{\Sigma}^{(\boldsymbol{\rho})}_{(\boldsymbol{n})}$ is positive definite for any $\boldsymbol{\rho}$ satisfying (5.6). Hence, $C^{(\boldsymbol{\rho})}_{(\boldsymbol{n})}$ in Model 5.2 is a valid Gaussian copula. One can see that the estimation of the parameters in (5.11) is involved with the calculation of multivariate Gaussian density function which depends on the length of the observer years. Let $\boldsymbol{b} = (b_1, \cdots, b_\tau)$ be vertices where each $b_j$ is equal to either $n_j$ or $n_j - 1$. Then the corresponding density function of the random vector of $\boldsymbol{Z}_{(\boldsymbol{\tau})}$ at $\boldsymbol{Z}_{(\boldsymbol{\tau})} = \boldsymbol{z}_{(\boldsymbol{\tau})}$ in (5.11) is given by

$$h(\boldsymbol{z}_{(\boldsymbol{\tau})}) = \sum \operatorname{sgn}(\boldsymbol{b}) \frac{\partial^{\bar{z}+\tau}}{\partial y_{1,1} \cdots \partial y_{1,n_1}, \partial y_{2,1} \cdots \partial y_{2,n_2}, \cdots, \partial y_{\tau,1}, \cdots \partial y_{\tau,n_\tau}} H\left(\boldsymbol{z}_{(\boldsymbol{\tau})}\right) \tag{5.12}$$

where the sum is taken over all vertices $\boldsymbol{b}$, and $\mathrm{sgn}(\boldsymbol{b})$ is given by

$$\mathrm{sgn}(\boldsymbol{b}) = \begin{cases} +1, & \text{if } b_j = n_j - 1 \text{ for an even number of } j\text{'s;} \\ \\ -1, & \text{if } b_j = n_j - 1 \text{ for an odd number of } j\text{'s.} \end{cases}$$

Here, we note that calculation of the density function in (5.12) can be difficult due to the following aspects of our model.

- Due to the discrete nature of the frequency observations, one can immediately check that the computational complexity in (5.12) grows exponentially with $\tau$.
- The calculation of each summation in (5.12), which requires a numerical multivariate integration due to the nature of multivariate Gaussian function, can be even difficult especially in high dimensions (Genz and Bretz, 2009)

However, here we avoid such difficulty by using the following copula representation which is inspired by factor copula representation in Krupskii and Joe (2013), Nikoloulopoulos and Joe (2015) and Kadhem and Nikoloulopoulos (2019). For $\boldsymbol{\rho}$ defined in (5.6) satisfying (5.9), we extend the modeling of $\boldsymbol{Z}_{(\boldsymbol{\tau})}$ by including the random effect $R$ so that the joint distribution of $\left(R, \boldsymbol{Z}_{(\boldsymbol{\tau})}\right)$ is given by

$$\begin{aligned} &H^*\left(r, \boldsymbol{z}_{(\boldsymbol{\tau})}\right) \\ &= C_{(\boldsymbol{n})}^{(\boldsymbol{\rho},\boldsymbol{\theta})}\left(\Phi(r), F_1(n_1), G_{1,1}(y_{1,1}), \cdots, G_{1,n_1}(y_{1,n_1}), \cdots F_\tau(n_\tau), G_{\tau,1}(y_{\tau,1}), \cdots, G_{\tau,n_\tau}(y_{1,n_\tau})\right). \end{aligned} \tag{5.13}$$

Naturally, by the property of the copula $C_{(\boldsymbol{n})}^{(\boldsymbol{\rho},\boldsymbol{\theta})}$, the joint distribution in (5.13) implies the joint distribution function in (5.11) in the following sense

$$H\left(\boldsymbol{z}_{(\boldsymbol{\tau})}\right) = \lim_{r\to\infty} H^*\left(r, \boldsymbol{z}_{(\boldsymbol{\tau})}\right),$$

which further implies that the random vector $\left(R, \boldsymbol{Z}_{(\boldsymbol{\tau})}\right)$ is a natural extension of the random vector $\boldsymbol{Z}_{(\boldsymbol{\tau})}$. Furthermore, reparameterization in (5.6) gives us a well-defined and natural dependence structure with the shared random effect $R$ so that claims across multiple years would be independent conditional on $R$. For example, if (5.6) holds and $\theta_1 = \theta_2 = 0$, then one can see that $\boldsymbol{Z}_{(\boldsymbol{\tau})}$ are not only conditionally independent but also marginally independent so that $\boldsymbol{Z}_t \perp \boldsymbol{Z}_{t'}$ for all $t \neq t'$. In addition to (5.6), if $\theta_3 = \theta_4 = 0$, then $\boldsymbol{M}$ is not only block-diagonal, but diagonal, which implies that frequency and severity are independent once the shared random effect $R$ is controlled. In other words, dependence between frequency and severity are fully explained by the shared random effect $R$. Finally, if (5.6) holds and $\theta_1 = \theta_2 = \theta_3 = \theta_4 = 0$, then $\boldsymbol{\Sigma}^{(\boldsymbol{\rho})}_{(\boldsymbol{n})}$ is diagonal, which implies our model specification includes the traditional model, which assumes independence among the claims in different years and independence between the frequency and severity.

The following theorem shows us the key idea of our copula representation where the observed claim $\boldsymbol{Z}_t$ for $t = 1, \ldots, \tau$ are independent conditional on the random effect $R$, and can be fitted into the special case of Model 5.1. In this regard, the copula in (5.13) has similar spirit as a factor copula. The corresponding copula of the distribution of $\boldsymbol{Z}_t$ conditional on $R$ is a Gaussian copula which can be represented as 1-factor copula. As a result, the distribution in (5.13) have 2-factor copula representation. However, such representation of the model increases the complexity of the notation while provides limited benefit in computational complexity, and hence we do not pursue such representation for the simplicity of the paper.

**Theorem 5.2.** *Suppose that $\boldsymbol{\rho}$ satisfies* (5.6) *and joint distribution function $H^*$ of the random vector $(R, \boldsymbol{Z}_t)$ is given by the factor copula model in* (5.13). *Then, we have the following results.*

*i. The distribution function of $\boldsymbol{Z}_{(\boldsymbol{\tau})}$ can be obtained as in* (5.11).

*ii. The density function of $\boldsymbol{Z}_{(\boldsymbol{\tau})} = \boldsymbol{z}_{(\boldsymbol{\tau})}$ conditional on $R = r$ is given by*

$$h^*\left(\boldsymbol{z}_{(\boldsymbol{\tau})}|r\right) = \prod_{t=1}^{\tau} h_t^*\left(\boldsymbol{z}_t|r\right)$$

*where $h_t^*(\cdot|r)$ is the conditional density function of $\boldsymbol{Z}_t$ conditional on $R = r$.*

*iii. $N_t \perp R$ for $t = 1, \ldots$ if and only if $\theta_1 = 0$.*

*iv. $\boldsymbol{y}_t \perp R$ for $t = 1, \ldots$ if and only if $\theta_2 = 0$.*

*Proof.* The proof of part i is trivial from the property of copula function. The proof of part ii, by the invariance property of the copula under the monotone transformation, we have that the corresponding copula $C$ of the conditional distribution of random vector $\boldsymbol{Z}_{(\boldsymbol{\tau})}$ conditional on $R = r$ is again a Gaussian copula. Furthermore, knowing that $C$ is a Gaussian copula, Theorem 5.1 shows that $\boldsymbol{Z}_1, \cdots, \boldsymbol{Z}_\tau$ are independent conditional on $R = r$. The proofs of part iii and iv are immediate from the property of Gaussian copulas. □

Based on this result in Theorem 5.2, one can obtain the joint density of $(\boldsymbol{Z}_1, \cdots, \boldsymbol{Z}_\tau)$ just with a single dimensional (numerical) integration as the following manner.

**Corollary 5.1.** *Consider the random vector $\boldsymbol{Z}_t$ under the settings in Model 5.2. Then, the joint density of $\boldsymbol{Z}_{(\boldsymbol{\tau})}$ is given as follows:*

$$h(\boldsymbol{z}_{(\boldsymbol{\tau})})$$
$$= \int \prod_{t=1}^{\tau} \left[ g_t^{[\text{joint}]*}\left(y_{t,1}, \cdots, y_{t,n_t}|r\right) \left( \Phi\left( \frac{\Phi^{-1}(F(n_t)) - \mu_t}{\sigma_t} \right) - \Phi\left( \frac{\Phi^{-1}(F(n_t - 1)) - \mu_t}{\sigma_t} \right) \right) \right] \phi(r)\mathrm{d}r$$

*where*

$$\mu_t = (\theta_1, \rho_1, \cdots, \rho_1) \left(\boldsymbol{\Sigma}_{tt}^{(\boldsymbol{\rho}^*)}\right)^{-1} \left(r, \Phi^{-1}\left(G(y_{t,1})\right), \cdots, \Phi^{-1}\left(G(y_{t,n_t})\right)\right)^{\mathrm{T}} \tag{5.14}$$

*and*

$$(\sigma_t)^2 = 1 - (\theta_1, \rho_1, \cdots, \rho_1) \left(\boldsymbol{\Sigma}_{tt}^{(\boldsymbol{\rho}^*)}\right)^{-1} (\theta_1, \rho_1, \cdots, \rho_1)^{\mathrm{T}} \tag{5.15}$$

*with $\rho_1^* = \theta_2$ and $\rho_2^* = \rho_2$. Here, $g_t^{[\text{joint}]*}(\cdot|r)$ is the density function of $\boldsymbol{Y}_t$ conditional on*

*$R = r$, and given by*

$$\begin{aligned} & g_t^{[\text{joint}]*}\left(y_{t,1}, \cdots, y_{t,n_t} | r\right) \\ & = \phi_{\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*}\left(\Phi^{-1}\left(G\left(y_{t,1}\right)\right), \cdots, \Phi^{-1}\left(G\left(y_{t,n_t}\right)\right)\right) \prod_{j=1}^{n_t} \frac{g(y_{t,j})}{\phi\left(\Phi^{-1}\left(G(y_{t,j})\right)\right)} \end{aligned}$$

*where*

$$\boldsymbol{\mu}^* = r\,\theta_2 \mathbf{1}_{n_t} \quad \textit{and} \quad \boldsymbol{\Sigma}^* = (1-\rho_2)\boldsymbol{I}_{n_t} + (\rho_2 - \theta_2^2)\boldsymbol{J}_{n_t \times n_t}.$$

*Proof of Corollary 5.1.* According to Theorem 5.2, we extend $\boldsymbol{Z}_{(\boldsymbol{\tau})}$ to the factor copula model $\left(R, \boldsymbol{Z}_{(\boldsymbol{\tau})}\right)$ having the distribution function in (5.13). Then, we have

$$\begin{aligned} h(\boldsymbol{z}_{(\boldsymbol{\tau})}) &= \int h^*(\boldsymbol{z}_{(\boldsymbol{\tau})} | r)\phi(r)\mathrm{d}r \\ &= \int \prod_{t=1}^{\tau} h_t^*(\boldsymbol{z}_t | r)\phi(r)\mathrm{d}r \\ &= \int \prod_{t=1}^{\tau} \left[g_t^{[\text{joint}]*}\left(y_{t,1}, \cdots, y_{t,n_t} | r\right) \mathbb{P}\left(N_t = n_t | y_{t,1}, \cdots, y_{t,n_t}, r\right)\right] \phi(r)\mathrm{d}r \end{aligned}$$

where $h^*(\cdot|r)$, and $h_t^*(\cdot|r)$ are the density functions of $\boldsymbol{Z}_{(\boldsymbol{\tau})}$ and $\boldsymbol{Z}_t$, respectively, and $g_t^{[\text{joint}]*}\left(y_{t,1}, \cdots, y_{t,n_t} | r\right)$ is the density function of $(Y_{t,1}, \cdots, Y_{t,n_t})$ at $(Y_{t,1}, \cdots, Y_{t,n_t}) = (y_{t,1}, \cdots, Y_{y,n_t})$ conditional on $R = r$. Here, the second equality is from Theorem 5.2, and the final equality is just conditional distribution expression of the joint density function.

Finally, it suffices to show that

$$\begin{aligned} \mathbb{P}\left(n_t | y_{t,1}, \cdots, y_{t,n_t}, r\right) &= \mathbb{P}(N_t \le n_t | r, y_{t,1}, \cdots, y_{t,n_t}) - \mathbb{P}(N_t \le n_t - 1 | r, y_{t,1}, \cdots, y_{t,n_t}) \\ &= \Phi\left(\frac{\Phi^{-1}(F(n_t)) - \mu_t}{\sigma_t}\right) - \Phi\left(\frac{\Phi^{-1}(F(n_t - 1)) - \mu_t}{\sigma_t}\right) \end{aligned}$$

and

$$g_t^{[\text{joint}]*}\left(y_{t,1}, \cdots, y_{t,n_t} | r\right) = \phi_{\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*}\left(\Phi^{-1}\left(G\left(y_{t,1}\right)\right), \cdots, \Phi^{-1}\left(G\left(y_{t,n_t}\right)\right)\right) \prod_{j=1}^{n_t} \frac{g(y_{t,j})}{\phi\left(\Phi^{-1}\left(G(y_{t,j})\right)\right)},$$

which are proved by Lemmas 6.1 and 6.2, respectively in Appendix C. □

**Remark 5.1.** *The model specification in Model 5.2, and equivalently Model 5.1, is not casual in the sense that the length or dimension of the observation varies depending on the value of the observation. For example, we have $\boldsymbol{z}_t = (n_t, y_{t,1})$ for $n_t = 1$ while $\boldsymbol{z}_t = (n_t, y_{t,1}, y_{t,2})$ for $n_t = 2$. Hence, the model itself does not seem to be well-defined as it is not even clear how to mathematically define cumulative distribution function or the joint density function. In Appendix E, we show how to interpret the density function and corresponding distribution function in Model 5.2 so that they are well-defined. Specifically, one can easily check that the copula function $C^{(\boldsymbol{\rho})}_{(\boldsymbol{n})}$ in Model 5.2 satisfies the inheritance property in the similar manner as in* (F.2)*, which further implies that Model 5.2 can be reformulated as Model 6.1 where the distribution and density functions are well-defined. Finally, one can easily show that the distribution and density functions in Model 5.2 are the same as those in Model 6.1 so that they are well-defined. We leave the detailed discussion in Appendix E. The discussion on the well-definedness of Model 5.2 but limited to one-year model can be also found in Oh et al. (2019).*

## 5.4 Simulation study

In this section, we conduct a simulation study to investigate the finite sample properties of the parameter estimates and effects of the dependences on them for the proposed method based on Model 5.2. We assume one risk class only, where the distribution function $F$ follows Poisson distribution with mean parameter $\lambda_0$ and the distribution function $G$ follows Weibull distribution with mean paramter $\xi_0$ and shape paramter $\nu$. Here, the parameters for the marginal part of severity are specified as $\xi_0 = \exp(8)$, and $\nu = 0.7$, which are the same for all scenarios. The portfolio of policyholders of size $I$ observed for three years ($\tau = 3$) are generated from the proposed model under 8 scenarios motivated by the real data analysis in Section 5.5. Table 5.1 provides the rest of parameter settings and the corresponding correlation coefficients.

Table 5.1: Parameter settings of the copula part for each scenario

| | | Parameter | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Scenario | I | $\lambda_0$ | $\theta_1$ | $\theta_2$ | $\theta_3$ | $\theta_4$ | $\rho_1$ | $\rho_2$ | $\rho_3$ | $\rho_4$ | $\rho_5$ |
| 1 | 500 | 2 | 0.3 | 0.3 | 0.5 | 0.5 | 0.34 | 0.34 | 0.09 | 0.09 | 0.09 |
| 2 | | | 0.3 | 0.3 | 0 | 0 | 0.09 | 0.09 | 0.09 | 0.09 | 0.09 |
| 3 | | | 0.3 | 0.7 | 0.5 | 0.5 | 0.46 | 0.74 | 0.09 | 0.21 | 0.49 |
| 4 | | | 0.3 | 0.7 | 0 | 0 | 0.21 | 0.49 | 0.09 | 0.21 | 0.49 |
| 5 | | | 0.7 | 0.3 | 0.5 | 0.5 | 0.46 | 0.34 | 0.49 | 0.21 | 0.09 |
| 6 | | | 0.7 | 0.3 | 0 | 0 | 0.21 | 0.09 | 0.49 | 0.21 | 0.09 |
| 7 | | | 0.7 | 0.7 | 0.5 | 0.5 | 0.74 | 0.74 | 0.49 | 0.49 | 0.49 |
| 8 | | | 0.7 | 0.7 | 0 | 0 | 0.49 | 0.49 | 0.49 | 0.49 | 0.49 |

For each scenario, Tables 5.2 and 5.3 summarize the simulation results from 500 independent Monte Carlo samples, including the relative bias and mean squared error (MSE) of the parameter estimates. Table 5.2 indicates that in all the scenarios, the estimates are close to the true parameters of the proposed model and shows that the relative bias and MSE are small.

Table 5.2: Relative bias in % for all the parameters from the each scenarios

| | RB (%) | | | | | | |
|---|---|---|---|---|---|---|---|
| Scenario | $\lambda_0$ | $\xi_0$ | $\nu$ | $\theta_1$ | $\theta_2$ | $\theta_3$ | $\theta_4$ |
| 1 | -0.21 | -0.06 | 0.17 | -1.48 | 1.08 | -0.06 | -0.27 |
| 2 | -0.26 | -0.26 | 1.19 | 0.84 | -0.98 | - | - |
| 3 | -0.52 | -0.03 | -10.77 | -9.81 | 1.16 | -1.19 | 0.10 |
| 4 | 0.05 | -0.53 | 1.74 | 6.09 | 0.64 | - | - |
| 5 | 0.00 | -0.08 | -0.29 | 0.33 | 1.43 | -1.36 | 0.14 |
| 6 | 0.38 | -0.31 | 18.92 | -1.70 | -0.63 | - | - |
| 7 | -0.16 | -0.03 | -20.82 | -19.90 | 0.54 | -0.63 | -0.01 |
| 8 | 0.14 | -0.50 | 13.95 | 9.68 | 1.00 | - | - |

## 5.5 Empirical application

In this section, we now calibrate the proposed model to a real auto insurance dataset, to examine dependence structure (i) between frequency and severity within a year, (ii) among two

Table 5.3: Mean absolute error for all the parameters from the 12 scenarios

| | MSE | | | | | | |
|---|---|---|---|---|---|---|---|
| Scenario | $\lambda_0$ | $\xi_0$ | $\nu$ | $\theta_1$ | $\theta_2$ | $\theta_3$ | $\theta_4$ |
| 1 | 0.0015 | 0.0008 | 0.0023 | 0.0012 | 0.0021 | 0.0007 | 0.0001 |
| 2 | 0.0010 | 0.0009 | 0.0011 | 0.0004 | 0.0001 | - | - |
| 3 | 0.0014 | 0.0009 | 0.0131 | 0.0039 | 0.0023 | 0.0008 | 0.0001 |
| 4 | 0.0015 | 0.0020 | 0.0023 | 0.0012 | 0.0001 | - | - |
| 5 | 0.0014 | 0.0011 | 0.0038 | 0.0015 | 0.0033 | 0.0008 | 0.0001 |
| 6 | 0.0015 | 0.0009 | 0.0023 | 0.0006 | 0.0001 | - | - |
| 7 | 0.0012 | 0.0011 | 0.0124 | 0.0064 | 0.0025 | 0.0007 | 0.0002 |
| 8 | 0.0017 | 0.0019 | 0.0069 | 0.0024 | 0.0001 | - | - |

distinct severities within a year, (iii) among frequencies across years, (iv) between frequency and severity in different years, and (v) between two severities in different years.

### 5.5.1 Data

For this empirical investigation, we employ a dataset from a general insurer in Singapore, which consists of a portfolio of personal automobile insurance policies with comprehensive coverages. The dataset has been obtained from General Insurance Association of Singapore, a trade association with representations from all the general insurance companies in Singapore. The claims experience observed from this dataset is longitudinal over a period of six years, from 1995 to 2000, and has 17,452 unique policyholders. Among the observations, we randomly sample 5000 policyholders. To calibrate the models, the observations for the first five years, 1995-1999 is used as in-sample, or training data, and in order to examine the performance of the model, we use the last year 2000 as the hold-out sample, or test data.

The dataset also contains a set of predictors that could further explain additional variation in the number of claims and the claim amounts. To summarize the variables observed, we have three categorical variables and one continuous variable: the gender with two levels (male and female), insured's age (Age) with four levelss including age 1 $\in (0, 25]$, age 2 $\in (25, 35]$, age 3 $\in (35, 65]$, and age 4 $\in (65, \infty]$, vehicle age (VehAge) with four levels including vehicle age 1 $\in [0, 1]$, vehicle age 2 $\in (1, 5]$, vehicle age 3 $\in (5, 10]$, and vehicle age 4 $\in (10, \infty]$, and vehicle's capacity expressed in log scale (logVehCapa).

Table 5.4 summarizes the description and simple statistics of these predictor variables which represent the risk characteristics of policyholders: Gender, Age, VehAge, and logVehCapa. In Singapore, as observed in this table, there is a disproportionate distribution by gender with more male than female drivers. When we the distribution of drivers by age, it is also not surprising to find fewer percentage of younger drivers, unlike that in other developed countries. The primary reason for this is the extremely expensive cost of owning and maintaining a car, in addition to the efficiency of the use of public transportation. During the period of observation, it is highly discourage to own a car for more than 10 years, and this reflected in this distribution.

Furthermore, a summary of the claim frequency over the years 1995 to 1999 is given in Table 5.5 and the average claim amount categorized by frequency and year is given in Table 5.6. This table suggests that the claims size appears to be unstable over time. We adjust the values of the individual severities, in order to satisfy that the average of individual severity over each year is the same as the average over the 2000 hold-out sample data with 4,659 observations.

### 5.5.2 Estimation result

For the data analysis, we consider the model with regression setting described in Corollary 5.1. We assume the distribution function, $F$, follows a Poisson distribution with mean parameter, $\lambda$, for the frequency, and the distribution function, $G$, follows a Weibull distribution with mean parameter, $\xi$, and shape parameter, $\nu$, for the severity component. With a log link function, we therefore have

$$\lambda = \exp(\boldsymbol{x}\boldsymbol{\beta}), \quad \text{and} \quad \xi = \exp(\boldsymbol{w}\boldsymbol{\gamma}),$$

Table 5.4: Observable policy characteristics used as covariates

| Categorical variables | Description | | | Proportions |
|---|---|---|---|---|
| Gender | Insured's sex: | | | |
| | Male = 1 | | | 80.03% |
| | Female = 0 | | | 19.97% |
| Age | The policyholder's issue age : | | | |
| | Age $\in (0, \ 25] = 1$ | | | 0.49% |
| | Age $\in (25, 35] = 2$ | | | 21.68% |
| | Age $\in (35, 65] = 3$ | | | 76.81% |
| | Age $\in (65, \infty] = 4$ | | | 1.03% |
| VehAge | Age of vehicle in years : | | | |
| | VehAge $\in [0, \ 1] = 1$ | | | 12.45% |
| | VehAge $\in (1, \ 5] = 2$ | | | 57.30% |
| | VehAge $\in (5, \ 10] = 3$ | | | 29.99% |
| | VehAge $\in (10, \infty] = 4$ | | | 0.25% |
| Continuous variables | | Min | Mean | Max |
| logVehCapa | Insured vehicle's capacity in cc | 6.49 | 7.19 | 8.82 |

Table 5.5: Number of observations by frequency and year

| | Train | | | | | | | Test | |
|---|---|---|---|---|---|---|---|---|---|
| Frequency | 1995 | 1996 | 1997 | 1998 | 1999 | Count | % of Total | 2000 | % of Total |
| 0 | 3103 | 3291 | 2501 | 2036 | 1751 | 12682 | 91.05 | 1360 | 92.39 |
| 1 | 232 | 212 | 266 | 214 | 219 | 1143 | 8.21 | 104 | 7.07 |
| 2 | 17 | 8 | 20 | 24 | 18 | 87 | 0.62 | 8 | 0.54 |
| 3 | 2 | 1 | 2 | 2 | 4 | 11 | 0.08 | 0 | 0 |
| Count | 3354 | 3512 | 2789 | 2276 | 1992 | 13923 | 100 | 4659 | 100 |

Table 5.6: Average severity by frequency and year

| | Train | | | | | | Test |
|---|---|---|---|---|---|---|---|
| Frequency | 1995 | 1996 | 1997 | 1998 | 1999 | Avg. severity | 2000 |
| 1 | 4742 | 4530 | 4567 | 5440 | 3895 | 4630 | 4557 |
| 2 | 6319 | 3633 | 3629 | 3781 | 3644 | 4200 | 2950 |
| 3 | 2630 | 1687 | 4991 | 6015 | 3065 | 3747 | - |
| 4 | - | - | - | - | - | - | - |
| Avg. severity | 4892 | 4431 | 4455 | 5156 | 3824 | 4553 | 4046 |

where $\boldsymbol{x}$ and $\boldsymbol{w}$ are the vectors of model matrices for each policyholder [1], and $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are the corresponding parameters for the frequency and severity, respectively. Hence, in this data

[1] In this example, $\boldsymbol{x}$ includes Gender, Age, and VehAge, and $\boldsymbol{w}$ includes VehCapa, Age, and VehAge.

analysis, we consider following parameters: $(\boldsymbol{\beta}, \boldsymbol{\gamma}, \nu, \theta_1, \theta_2, \theta_3, \theta_4)$.

Table 5.7 presents the summary statistics for the model estimation results. This table provides details of the estimated parameters for the frequency part, the severity part, as well as the copula part. There are four measures detailed in this table: estimates (est), standard errors (std.error), t statistics (t), and corresponding p-values. Note that the asterisk sign (*) indicates that the estimate is significant at 0.05 level. From the table, the results are as expected. For example, despite the disproportionate percentage, male drivers are expected to incur more accidents than female drivers. When analyzed by age, broadly speaking, both frequency and severity tend to decrease with age. Elderly drivers, for example, have fewer number of accidents with smaller average costs per accident than drivers less than 25 years old.

In terms of understanding the presence of dependence, Table 5.7 also summarizes estimates of the four copula parameters of dependence as described by $\theta_i$, for $i = 1, 2, 3, 4$, and the estimates are not all significantly nonzero at the 5% level. For the interpretation of copula parameters in Table 5.7, one can recall the following meaning of $\theta_1, \theta_2, \theta_3$, and $\theta_4$:

- $\theta_1$: dependence parameter between the common random effect $R$ and the frequency for every year,
- $\theta_2$: dependence parameter between the common random effect $R$ and each severity for every year,
- $\theta_3$, $\theta_4$: dependence parameters between a frequency and each severity within a year not explained by $R$.

Thus, according to the estimation results which shows that only $\theta_1$ and $\theta_2$ are significantly different from zero, we can claim presence of both types of dependence; temporal dependence of claim frequencies and severities as well as dependence between the frequency and severity can be explained by common random effect $R$. On the other hand, there is weak evidence of

Table 5.7: Estimation result

| parameter | est | std.error | t | p-value | |
|---|---|---|---|---|---|
| **Frequency part** | | | | | |
| (Intercept) | -2.237 | 0.289 | -7.749 | <.0001 | * |
| Gender | 0.125 | 0.067 | 1.865 | 0.0623 | |
| VehAge2 | 0.048 | 0.101 | 0.477 | 0.6336 | |
| VehAge3 | -0.146 | 0.109 | -1.339 | 0.1806 | |
| VehAge4 | 0.835 | 0.443 | 1.886 | 0.0594 | |
| Age2 | 0.323 | 0.270 | 1.197 | 0.2313 | |
| Age3 | 0.156 | 0.268 | 0.583 | 0.5601 | |
| Age4 | -0.569 | 0.460 | -1.237 | 0.2161 | |
| **Severity part** | | | | | |
| (Intercept) | 3.889 | 0.938 | 4.148 | <.0001 | * |
| logVehCapa | 0.700 | 0.108 | 6.468 | <.0001 | * |
| VehAge2 | -0.010 | 0.097 | -0.099 | 0.9212 | |
| VehAge3 | -0.060 | 0.110 | -0.547 | 0.5843 | |
| VehAge4 | -0.624 | 0.565 | -1.106 | 0.2690 | |
| Age2 | -1.092 | 0.478 | -2.284 | 0.0224 | * |
| Age3 | -0.976 | 0.475 | -2.055 | 0.0400 | * |
| Age4 | -0.969 | 0.668 | -1.451 | 0.1468 | |
| $\nu$ | 0.802 | 0.045 | 17.910 | <.0001 | * |
| **Copula part** | | | | | |
| $\theta_1$ | 0.263 | 0.048 | 5.509 | <.0001 | * |
| $\theta_2$ | 0.057 | 0.071 | 0.795 | 0.4266 | |
| $\theta_3$ | 0.409 | 0.138 | 2.967 | 0.0030 | * |
| $\theta_4$ | 0.445 | 0.133 | 3.341 | 0.0008 | * |

dependence between a frequency and each frequency within a year not explained by $R$.

While the values of $\theta$ tells us the relationship between the common random effects and claims, one can directly quantify the magnitude of dependence among the claims by observing the estimated values of $\rho$'s. According to the model specification in (5.6), the estimates of dependence parameters, $\rho_1$, $\rho_2$, $\rho_3$, $\rho_4$, and $\rho_5$ are calculated from the estimates of $\theta_1$, $\theta_2$, $\theta_3$, and $\theta_4$ as in (5.6), and their standard errors are obtained using delta method. Table 5.8 summarizes the derived estimates together with the resulting standard errors of $\rho_1$, $\rho_2$, $\rho_3$, $\rho_4$, and $\rho_5$.

It is interesting to observe that there is now a clearer evidence of all types of dependencies in

Table 5.8: Derived estimates and standard errors of $\rho$'s

| parameter | est | std.error | t | p-value | |
|---|---|---|---|---|---|
| $\rho_1$ | 0.1968 | 0.074 | 2.655 | 0.0079 | * |
| $\rho_2$ | 0.2015 | 0.119 | 1.688 | 0.0915 | |
| $\rho_3$ | 0.0690 | 0.025 | 2.754 | 0.0059 | * |
| $\rho_4$ | 0.0149 | 0.019 | 0.780 | 0.4355 | |
| $\rho_5$ | 0.0032 | 0.008 | 0.398 | 0.6909 | |

our multi-year microlevel model. For example, $\rho_1$ describes correlation between a frequency and a severity within a single year, and results provide strong evidence of a positive dependence. The estimate for $\rho_1$ is 0.1968 with a standard error of 0.074, which leads to a very small p-value indicating significantly different from zero. Using the results from (5.6), despite the non-significance of $\theta_3$ and $\theta_4$ directly drawn from the estimated model, there is a clear inherent dependence driven by the shared random effect through the interplay with $\theta_1$ and $\theta_2$. A similar argument can be said of the other $\rho$'s.

### 5.5.3 Validation

For validation of the proposed model in terms of the individual loss prediction for the 1,472 policyholders in the hold-out sample, we compare the following four models: full model, nested model 1 with $\theta_3 = \theta_4 = 0$, the nested model 2 with $\theta_1 = \theta_2 = 0$, and the nested model 3 with $\theta_1 = \theta_2 = \theta_3 = \theta_4 = 0$. We measure the quality of prediction as mean squared error (MSE) of average of individual loss prediction over 5,000 Monte Carlo simulations under the estimation result from each model. We also use other measures such as root-mean-square deviation (RMSE), mean absolute error (MAE), and the Gini index in Frees et al. (2011b, 2014c). For example, MSE for full model is given as

$$\widehat{\mathrm{MSE}}^{[\mathrm{full}]} = \frac{1}{1472} \sum_{i=1}^{1472} \left( S_i - \hat{S}_i^{[\mathrm{full}]} \right)^2,$$

where $S_i$ is the observed aggregate loss for the $i$-th policyholder in 2000, and $\hat{S}_i^{[\mathrm{full}]}$ is the average of predicted aggregate loss for the $i$-th policyholder over 5,000 MC samples from based on the full model. The results are shown in Table 5.9. In the table, full model shows the best performance in terms of MSE and RMSE, and nested model 1 shows best performance in terms of MAE. On the other hand, nested model 2 shows the best performance in terms of the Gini index.

Table 5.9: Means squared error

| | Full | Nested 1 | Nested 2 | Nested 3 |
|---|---|---|---|---|
| RMSE | 2445.409 | 2448.719 | 2445.519 | 2448.276 |
| MSE | 5980026 | 5996227 | 5980564 | 5994053 |
| MAE | 596.87 | 524.2761 | 605.0203 | 530.9766 |
| Gini | 28.535 | 30.478 | 27.560 | 29.547 |

## 5.6 Final remarks

This article focuses on the development of a multi-year microlevel collective risk model which accounts for a flexible dependence structure for claim frequencies and claim severities. The common theme in the literature is a framework that regards dependence between claim frequency and the average severity. Our motivating example demonstrates that for these types of dependence models, the copula structure can be constrained. Here, we also show that it is even difficult to arrive at the naive assumption of independence among severities.

In our multi-year microlevel collective risk model, we develop a shared random effects framework that captures various relevant types of dependence between claim frequencies and claim severities over multiple years. The shared random effect parameter induces several forms of dependence; it has similar structure to a one-factor copula model previously studied. Our proposed scheme has the advantages of not only ease of computation but also the capacity to draw intuitive interpretation to the results. Furthermore, it covers other types of dependent frequency and severity models that have previously been studied. One can see that both one-year dependent compound risk model and traditional independent compound risk model are special cases of our proposed model, where $\theta_1 = \theta_2 = 0$ and $\theta_1 = \theta_2 = \theta_3 = \theta_4 = 0$, respectively. In the paper, we additionally provide an efficient way to obtain the joint density of multi-year claim required without heavy numerical integration.

We calibrated our proposed with a dataset from a Singapore automobile insurance company, which contains policy characteristics and microlevel claims information for multiple years. The estimation results show us that all five types of correlations considered in a multi-year microlevel collective risk model are statistically significant. We note that the driving force for the dependencies originates from the shared random effect parameter. On top of that, out-of-sample validation results with the proposed model show us that it can be helpful to consider various types of dependence to increase the prediction performances.

# Chapter 6

# Conclusion with future works

Throughout this thesis, we explore applications of random effects models in actuarial ratemaking, considering one usual characteristic of P&C insurance datasets, longitudinality. In conclusion, key findings from the previous chapters are listed first. Finally, possible directions of future research are suggested.

## 6.1 Major contributions

In Chapter 2, the concept of Bregman divergence is explored, which has some good properties for statistical modeling and can be connected to diverse model selection diagnostics as in Goh and Dey (2014). We can apply model diagnostics derived from the Bregman divergence for testing robustness of priors both on the naive model, which assumes that random effect has point mass as its prior density, and the proposed model, which assumes a continuous prior density of random effect. This approach provides insurance companies concrete framework for testing the presence of random effects in both claim frequency and severity and furthermore appropriate hierarchical model which can explain both observed and unobserved heterogeneity of the policyholders for insurance ratemaking.

Chapter 3 explores the benefits of using random effects for predicting insurance claims observed longitudinally, or over a period of time, within a two-part framework relaxing the assumption of independence. More specifically, we introduce a generalized formula for credibility premium of compound sum with dependence. This extends and integrates previous work in both credibility premium of compound sums and dependent two-part compound risk models. In this generalized formula of credibility premium of compound sum, one can derive a dependence function, $D_N(\gamma)$, that offers an informative measure of the strength and direction of the association between frequency and severity. This function is easy to interpret and allows for practical implementation useful for actuarial ratemaking.

In Chapter 4, Bayesian credibility premium is formulated under a change of probability measure within the copula framework. Such reformulation is demonstrated using the multivariate generalized beta of the second kind (GB2) distribution. Within this family of GB2 copulas, it is possible to derive explicit form of Bayesian credibility premium even if the marginal distribution does not follow univariate GB2 distribution.

Chapter 5 introduces a shared random effects framework that captures various types of dependence between claim frequencies and claim severities over multiple years. It is a clear extension of earlier works on one-year dependent frequency-severity models and on random effects model for capturing serial dependence of claims. We develop not only a general framework but also concrete examples of model specification using a family of elliptical copulas.

## 6.2 Possible directions of future research

For additional furture works, one can consider shared random effects models for multi-peril claims. Usually, a P&C insurance policy consists of various types of claim, for example, such as collision, bodily injury liability, or property damage liability in the case of automobile insurance. In the presence of available covariates, impact of covariates for each type of claim could be different so that one needs to apply different estimation coefficient for each claim type. Besides, it is natural to expect that various types of claims have different correlation with the unobserved heterogeneity in risk, such as driving habits. For example, at-fault liability claim is usually at the control of the driver so that the occurrence of such claim is strongly related with driving habits. However, for glass damage claim, it is almost out of control of the driver (for example, due to heavy hailstorm) so that its association with driving habits may be quite low. In the end, a company needs to develop a predictive model which utilizes the covariate information to calculate 'a priori' premium as well as bonus-malus factor which utilizes past claim history of each policyholder. To handle such issue, one can propose a model where the shared random effects represent underlying unobserved heterogeneity, which affects the claims from all coverages simultaneously so that the shared random effects induce natural dependence among the claims from multiple coverages and over time.

To elaborate this idea with more details, let us define $N_{it}^{(j)}$ as number of claim type $j \in \{1, 2, \ldots, J\}$, for $i$th policyholder in year $t$. Here the numbers of claims are affected by both observable covariates and associated regression coefficient as well as (common) unobserved heterogeneity factor $\theta_i$, which is shared for all types and every years of claim of policyholder $i$, which can be postulated as follows:

$$w^{(j)} N_{it}^{(j)} | \mathbf{x}_{it}, e_{it}, \theta_i \overset{indep}{\sim} \mathcal{P}(\theta_i w^{(j)} \nu_{it}^{(j)}) \text{ and } \mathbb{E}\left[\theta_i\right] = 1,$$

where $\nu_{it}^{(j)} = \exp(\mathbf{x}_{it}\alpha^{(j)})$ and $w^{(j)}$ is (unknown) weight for each $j^{th}$ line of business. Here $\nu_{it}^{(j)}$ accounts for the observed heterogeneity in risk of policyholder $i$ at time $t$ for coverage $j$ while multiplicative random effect $\theta_i$ accounts for the unobserved heterogeneity in risk of policyholder $i$. Note that $w^{(j)}$ captures magnitude of (possible) overdispersion for $j^{th}$ line of business so that the dependence structure among multiple coverages over time can be modeled in a flexible manner.

Further, it can also be of interest to construct dynamic random effects models for capturing evolution of the unobserved heterogeneity. Throughout this thesis, the random effects which accout for the unobserved heterogeneity have been assumed to be static. In other words, they are unknown but fixed, which might not be suitable to describe possible evolution of the unobserved heterogeneity, such as driving habits or experiences. In this regard, one can suggest the following model specification to describe the average severity $C_{it}$:

$$\ln C_{it} = \mathbf{x}_{it}\beta + \frac{1}{\sqrt{t}}\epsilon_i(t) + \sigma Z_i.$$

Here $Z_i \sim \mathcal{N}(0, 1)$ and captures the static portion of unobserved heterogeneity and $\epsilon_i(t)$ is a standard brownian motion and captures dynamic portion of unobserved heterogeneity in partial because use of brownian motion in variance structure induces stronger correlation between the current claims and recent claims than the outdated claims. We laid sufficient foundation to conduct these future research works.

# Appendix A. Proof of Theorem 3.1 and Corollary 3.2

In this appendix, we provide the details of the derivation for the expression of the predictive mean of the aggregate claim as defined by $S_{T+1} = N_{T+1}\overline{C}_{T+1}$ according to our random effects model specification. For simplicity, here we drop the subscript $i$ and the conditioning argument on $\mathbf{x}$ so that

$$\mathbb{E}\left[S_{T+1}|\mathbf{n}_T, \overline{\mathbf{c}}_T\right] = \mathbb{E}\left[N_{T+1}\overline{C}_{T+1}|\mathbf{n}_T, \overline{\mathbf{c}}_T\right] = \mathbb{E}\left[N_{T+1}\mathbb{E}\left[\overline{C}_{T+1}|N_{T+1}\right]|\mathbf{n}_T, \overline{\mathbf{c}}_T\right].$$

In all cases of the average severity models, the conditional mean of $\overline{C}_t$ is given as

$$\mathbb{E}\left[\overline{C}_t|N_t\right] = \theta^C e^{\mathbf{x}_t\beta}\, e^{N_t\gamma} = \theta^C \tilde{\mu}_t\, e^{N_t\gamma}.$$

Therefore, predictive mean of $S_{T+1}$ can be expressed as follows under our severity model specifications:

$$\begin{aligned}
\mathbb{E}\left[S_{T+1}|\mathbf{n}_T, \overline{\mathbf{c}}_T\right] &= \mathbb{E}\left[N_{T+1}\overline{C}_{T+1}|\mathbf{n}_T, \overline{\mathbf{c}}_T\right] = \mathbb{E}\left[N_{T+1}\mathbb{E}\left[\overline{C}_{T+1}|N_{T+1}\right]|\mathbf{n}_T, \overline{\mathbf{c}}_T\right] \\
&= \mathbb{E}\left[N_{T+1}\theta^C\tilde{\mu}_{T+1}\, e^{N_{T+1}\gamma}|\mathbf{n}_T, \overline{\mathbf{c}}_T\right] \\
&= \mathbb{E}\left[\theta^C\tilde{\mu}_{T+1}|\mathbf{n}_T, \overline{\mathbf{c}}_T\right] \times \mathbb{E}\left[N_{T+1}e^{N_{T+1}\gamma}|\mathbf{n}_T\right] \\
&= \mathbb{E}\left[\theta^C|\mathbf{n}_T, \overline{\mathbf{c}}_T\right]\tilde{\mu}_{T+1} \times \mathbb{E}\left[M'_{N_{T+1}|\mathbf{n}_T}(\gamma)\right] \\
&= \mathbb{E}\left[\theta^C|\mathbf{n}_T, \overline{\mathbf{c}}_T\right]\tilde{\mu}_{T+1} \times \frac{r_T}{\tilde{r}_T}\nu_{T+1}e^{\gamma}\left[1-\left(\frac{\nu_{T+1}}{\tilde{r}_T}\right)(e^{\gamma}-1)\right]^{-r_T-1} \\
&= \mathbb{E}\left[\theta^C|\mathbf{n}_T, \overline{\mathbf{c}}_T\right]\tilde{\mu}_{T+1} \times \frac{r+\sum_{t=1}^T n_t}{r+\sum_{t=1}^T \nu_t}\nu_{T+1} \times e^{\gamma}\left[1-\left(\frac{\nu_{T+1}}{\tilde{r}_T}\right)(e^{\gamma}-1)\right]^{-r_T-1},
\end{aligned}$$

where we have used the following results which can be immediately deduced: $M_{N_{T+1}|\mathbf{n}_T}(z) = [1-(\nu_{T+1}/\tilde{r}_T)(e^z-1)]^{-r_T}$ and $\mathbb{E}\left[N_{T+1}e^{\gamma N_{T+1}}|\mathbf{n}_T\right] = M'_{N_{T+1}|\mathbf{n}_T}(\gamma)$.

Note that if $\gamma = 0$, then predictive mean of $S_{T+1}$ is reduced to the product of predictive means of $N_{T+1}$ and $\overline{C}_{T+1}$ as follows:

$$\begin{aligned}
\mathbb{E}\left[S_{T+1}|\mathbf{n}_T, \overline{\mathbf{c}}_T\right] &= \mathbb{E}\left[\theta^C|\mathbf{n}_T, \overline{\mathbf{c}}_T\right] \tilde{\mu}_{T+1} \times \frac{r+\sum_{t=1}^T n_t}{r+\sum_{t=1}^T \nu_t}\nu_{T+1} \\
&= \mathbb{E}\left[\theta^C|\mathbf{n}_T, \overline{\mathbf{c}}_T\right] \mu_{T+1} \times \frac{r+\sum_{t=1}^T n_t}{r+\sum_{t=1}^T \nu_t}\nu_{T+1} = \mathbb{E}\left[\overline{C}_{T+1}|\mathbf{n}_T, \overline{\mathbf{c}}_T\right] \times \mathbb{E}\left[N_{T+1}|\mathbf{n}_T\right].
\end{aligned}$$

This is because $\mu_{T+1} = \exp(\mathbf{x}_{T+1}\beta + N_{T+1}\gamma) = \exp(\mathbf{x}_{T+1}\beta + N_{T+1}\cdot 0) = \exp(\mathbf{x}_{T+1}\beta) = \tilde{\mu}_{T+1}$.

Since $\theta^C = 1$ under GB2 model, we have that

$$\begin{aligned}
\mathbb{E}\left[S_{T+1}|\mathbf{n}_T, \overline{\mathbf{c}}_T\right] &= \mathbb{E}\left[\theta^C|\mathbf{n}_T, \overline{\mathbf{c}}_T\right] \tilde{\mu}_{T+1} \times \frac{r+\sum_{t=1}^T n_t}{r+\sum_{t=1}^T \nu_t}\nu_{T+1} \times e^{\gamma}\left[1-\left(\frac{\nu_{T+1}}{\tilde{r}_T}\right)(e^{\gamma}-1)\right]^{-r_T-1} \\
&= \tilde{\mu}_{T+1} \times \frac{r+\sum_{t=1}^T n_t}{r+\sum_{t=1}^T \nu_t}\nu_{T+1} \times e^{\gamma}\left[1-\left(\frac{\nu_{T+1}}{\tilde{r}_T}\right)(e^{\gamma}-1)\right]^{-r_T-1}.
\end{aligned}$$

Under Gamma GLMM, $\mathbb{E}\left[\theta^C|\mathbf{n}_T, \overline{\mathbf{c}}_T\right]$ has no closed form but needs to be numerically evaluated, thereby losing computational ease, since we need to calculate $\mathbb{E}\left[S_{T+1}|\mathbf{n}_T, \overline{\mathbf{c}}_T\right]$ for each policyholder. For numerical integration, the following identity has been used:

$$\mathbb{E}\left[\theta^C|\mathbf{n}_T, \overline{\mathbf{c}}_T\right] = \int \theta\, \pi_C(\theta|\mathbf{n}_T, \overline{\mathbf{c}}_T)d\theta = \frac{\int \theta\, f(\overline{\mathbf{c}}_T|\mathbf{n}_T, \theta)\pi_C(\theta)d\theta}{\int f(\overline{\mathbf{c}}_T|\mathbf{n}_T, \theta)\pi_C(\theta)d\theta} \approx \frac{\sum_{j=1}^J \hat{\theta}_j\, f(\overline{\mathbf{c}}_T|\mathbf{n}_T, \hat{\theta}_j)}{\sum_{j=1}^J f(\overline{\mathbf{c}}_T|\mathbf{n}_T, \hat{\theta}_j)},$$

where $\hat{\theta}_j$'s are generated from $\mathcal{LN}\left(-\frac{\sigma^2}{2}, \sigma^2\right)$.

Under MVGP, we know that $\theta^C|\mathbf{n}_T, \overline{\mathbf{c}}_T \sim \mathcal{IG}(k_T+1, w_T)$ where $w_T = k + \sum_{t=1}^T S_t/\phi\mu_t$ and $k_T = k + \sum_{t=1}^T n_t/\phi$ because

$$\begin{aligned}
\pi_C(\theta|\mathbf{n}_T, \overline{\mathbf{c}}_T) &\propto \pi_C(\theta)\prod_{t=1}^T f(\overline{c}_t|n_t) \propto \theta^{-k-2}\exp\left(-\frac{k}{\theta}\right)\prod_{t=1}^T \theta^{\frac{n_t}{\phi}}\exp\left(-\frac{S_t/\phi\mu_t}{\theta}\right) \\
&\propto \theta^{-(\sum_{t=1}^T n_t/\phi+k+1)-1}\exp\left(-\frac{1}{\theta}(k+\sum_{t=1}^T S_t/\phi\mu_t)\right).
\end{aligned}$$

Hence, $\mathbb{E}\left[\theta^C|\mathbf{n}_T, \bar{\mathbf{c}}_T\right] = \dfrac{w_T}{k_T} = \dfrac{k + \sum_{t=1}^T S_t/\phi\mu_t}{k + \sum_{t=1}^T n_t/\phi} = \dfrac{k\phi + \sum_{t=1}^T S_t/\mu_t}{k\phi + \sum_{t=1}^T n_t}$. Therefore, we have

$$
\begin{aligned}
\mathbb{E}\left[S_{T+1}|\mathbf{n}_T, \bar{\mathbf{c}}_T\right] &= \mathbb{E}\left[\theta^C|\mathbf{n}_T, \bar{\mathbf{c}}_T\right] \tilde{\mu}_{T+1} \times \frac{r + \sum_{t=1}^T n_t}{r + \sum_{t=1}^T \nu_t}\nu_{T+1} \times e^{\gamma}\left[1 - \left(\frac{\nu_{T+1}}{\tilde{r}_T}\right)(e^{\gamma} - 1)\right]^{-r_T - 1} \\
&= \frac{k\phi + \sum_{t=1}^T S_t/\mu_t}{k\phi + \sum_{t=1}^T n_t}\tilde{\mu}_{T+1} \times \frac{r + \sum_{t=1}^T n_t}{r + \sum_{t=1}^T \nu_t}\nu_{T+1} \times e^{\gamma}\left[1 - \left(\frac{\nu_{T+1}}{\tilde{r}_T}\right)(e^{\gamma} - 1)\right]^{-r_T - 1}.
\end{aligned}
$$

Finally, under MVGB2, we know that $\theta^C|\mathbf{n}_T, \bar{\mathbf{c}}_T \sim \mathcal{GIG}(k_T + 1, w^*_{T,p}, p)$ where $z_t = \dfrac{\Gamma(n_t/\phi + 1/p)}{\Gamma(n_t/\phi)}$, $w = \dfrac{\Gamma(k+1)}{\Gamma(k + 1 - 1/p)}$, $w^*_{T,p} = [w^p + \sum_{t=1}^T (z_t \bar{c}_t \mu_t^{-1})^p]^{1/p}$, $k_T = k + \sum_{t=1}^T n_t/\phi$ because

$$
\begin{aligned}
\pi_C(\theta|\mathbf{n}_T, \bar{\mathbf{c}}_T) \propto \pi_C(\theta)\prod_{t=1}^T f(\bar{c}_t|n_t) &\propto \theta^{-p(k+1)-1}\exp\left(-\frac{w^p}{\theta^p}\right)\prod_{t=1}^T \theta^{-pn_t/\phi}\exp\left(-\left(\frac{\bar{c}_t z_t/\mu_t}{\theta}\right)^p\right) \\
&\propto \theta^{-p(\sum_{t=1}^T n_t/\phi + k + 1) - 1}\exp\left(-\frac{1}{\theta^p}\left[w^p + \sum_{t=1}^T (\bar{c}_t z_t/\mu_t)^p\right]\right).
\end{aligned}
$$

This leads to the following posterior mean of $\theta^C$:

$$
\begin{aligned}
\mathbb{E}\left[\theta^C|\mathbf{n}_T, \bar{\mathbf{c}}_T\right] = w^*_{T,p}\frac{\Gamma(k_T + 1 - 1/p)}{\Gamma(k_T + 1)} &= \left(w^p + \sum_{t=1}^T (z_t \bar{c}_t \mu_t^{-1})^p\right)^{1/p}\frac{\Gamma(k_T + 1 - 1/p)}{\Gamma(k_T + 1)} \\
&= \left(w^p + \sum_{t=1}^T\left(\frac{\bar{c}_t}{\mu_t}\frac{\Gamma(n_t/\phi + 1/p)}{\Gamma(n_t/\phi)}\right)^p\right)^{1/p}\frac{\Gamma(k_T + 1 - 1/p)}{\Gamma(k_T + 1)} \\
&= \left(w^p + \sum_{t=1}^T\left(\frac{\bar{c}_t n_t}{\mu_t\phi}\frac{\Gamma(n_t/\phi + 1/p)}{\Gamma(n_t/\phi + 1)}\right)^p\right)^{1/p}\frac{\Gamma(k_T + 1 - 1/p)}{\Gamma(k_T + 1)} \\
&= \left(\sum_{t=1}^T\left(\frac{S_t}{\mu_t}\frac{\Gamma(n_t/\phi + 1/p)}{\phi\Gamma(n_t/\phi + 1)}\right)^p + w^p\right)^{1/p}\frac{\Gamma(k_T + 1 - 1/p)}{\Gamma(k_T + 1)}.
\end{aligned}
$$

Finally, we have

$$\begin{aligned}
\mathbb{E}\left[S_{T+1}|\mathbf{n}_T, \overline{\mathbf{c}}_T\right] &= \mathbb{E}\left[\theta^C|\mathbf{n}_T, \overline{\mathbf{c}}_T\right]\, \tilde{\mu}_{T+1} \times \frac{r + \sum_{t=1}^T n_t}{r + \sum_{t=1}^T \nu_t}\nu_{T+1} \, \times e^\gamma \left[1 - \left(\frac{\nu_{T+1}}{\tilde{r}_T}\right)(e^\gamma - 1)\right]^{-r_T - 1} \\
&= \left(\sum_{t=1}^T \left(\frac{S_t}{\mu_t}\, \frac{\Gamma(n_t/\phi + 1/p)}{\phi\Gamma(n_t/\phi + 1)}\right)^p + w^p\right)^{1/p} \frac{\Gamma(k_T + 1 - 1/p)}{\Gamma(k_T + 1)}\tilde{\mu}_{T+1}\times \\
&\quad \frac{r + \sum_{t=1}^T n_t}{r + \sum_{t=1}^T \nu_t}\nu_{T+1} \, \times e^\gamma \left[1 - \left(\frac{\nu_{T+1}}{\tilde{r}_T}\right)(e^\gamma - 1)\right]^{-r_T - 1}.
\end{aligned}$$

# Appendix B. The derivation of GP and GB2 distributions

Here we suppress all the subscripts and let $\mu = e^{\mathbf{x}\beta + n\gamma}$. We see that when $\overline{C}|n, \theta^C \sim \mathcal{G}\left(\frac{n}{\phi}, \theta^C \mu \frac{\phi}{n}\right)$, the density of $\overline{C}|n, \theta^C$ is given as

$$f_{\overline{C}|N,\theta^C}(\overline{c}|n,\theta) = \frac{1}{\Gamma(n/\phi)} \left(\frac{n}{\theta\mu\phi}\right)^{n/\phi} \overline{c}^{n/\phi-1} \exp\left(-\frac{n\overline{c}}{\theta\mu\phi}\right).$$

For the random effects, when $\theta^C \sim \mathcal{IG}(k+1, k)$, the density of $\theta^C$ is given as

$$\pi_C(\theta) = \frac{1}{\Gamma(k+1)} \left(\frac{k}{\theta}\right)^{k+1} \exp\left(-\frac{k}{\theta}\right)\frac{1}{\theta}.$$

By integrating out the random effect, we can show that $\overline{C}|n \sim \mathcal{GP}\left(k+1, \mu k \frac{\phi}{n}, \frac{n}{\phi}\right)$ as follows:

$$\begin{aligned}
f(\overline{c}|n) &= \int_0^\infty f(\overline{c}|n,\theta) g(\theta) d\theta \\
&= \int_0^\infty f(\overline{c}|n,\theta) \frac{1}{\Gamma(k+1)} \left(\frac{k}{\theta}\right)^{k+1} \exp\left(-\frac{k}{\theta}\right)\frac{1}{\theta} d\theta \\
&= \frac{\overline{c}^{n/\phi-1}(n/\mu\phi)^{n/\phi} k^{k+1}}{\Gamma(n/\phi)\Gamma(k+1)} \int_0^\infty \theta^{-k-n/\phi-2} \exp\left(-\frac{k + n\overline{c}/\mu\phi}{\theta}\right) d\theta \\
&= \frac{\overline{c}^{n/\phi-1}(n/\mu\phi)^{n/\phi} k^{k+1}}{\Gamma(n/\phi)\Gamma(k+1)} \frac{\Gamma(n/\phi + k + 1)}{(k + n\overline{c}/\mu\phi)^{n/\phi+k+1}} \\
&= \frac{\Gamma(n/\phi + k + 1)}{\Gamma(n/\phi)\Gamma(k+1)} \frac{\overline{c}^{-1}(n\overline{c}/\mu\phi)^{n/\phi} k^{k+1}}{(k + n\overline{c}/\mu\phi)^{n/\phi+k+1}} \\
&= \frac{\Gamma(n/\phi + k + 1)}{\Gamma(n/\phi)\Gamma(k+1)} \frac{\overline{c}^{n/\phi-1}(k\mu\phi/n)^{k+1}}{(\overline{c} + k\mu\phi/n)^{n/\phi+k+1}}.
\end{aligned}$$

Note that if $Y \sim \mathcal{GP}(a, \xi, \tau)$, then the density function is given as follows:

$$f(y|a,\xi,\tau) = \frac{\Gamma(a+\tau)}{\Gamma(a)\Gamma(\tau)} \frac{\xi^a y^{\tau-1}}{(y+\xi)^{a+\tau}}. \tag{B.1}$$

We know that $\theta^C|\mathbf{n}_T, \overline{\mathbf{c}}_T \sim \mathcal{IG}(k_T+1, w_T)$ where $w_T = k+\sum_{t=1}^T S_t/\phi\mu_t$ and $k_T = k+\sum_{t=1}^T n_t/\phi$ from Appendix A. Thus, based on the derivation of the marginal GP density, it can be shown that $\overline{C}_{T+1}|\mathbf{n}_{T+1}, \overline{\mathbf{c}}_T \sim \mathcal{GP}\left(k_T+1, \mu_{T+1}w_T\dfrac{\phi}{n_{T+1}}, \dfrac{n_{T+1}}{\phi}\right)$ and

$$\mathbb{E}\left[\overline{C}_{T+1}|\mathbf{n}_{T+1}, \overline{\mathbf{c}}_T\right] = \frac{w_T}{k_T}\mu_{T+1} = \frac{k\phi + \sum_{t=1}^T S_t/\mu_t}{k\phi + \sum_{t=1}^T n_t}\mu_{T+1}.$$

Likewise, let us assume that $\overline{C}|n, \theta^C \sim \mathcal{GG}(n/\phi, \theta^C\mu/z, p)$ with the following density

$$f(\overline{c}|n, \theta) = \frac{p}{\Gamma(n/\phi)}\left(\frac{z}{\theta\mu}\right)^{pv}\overline{c}^{pv-1}\exp\left(-\left(\frac{\overline{c}z}{\theta\mu}\right)^p\right),$$

where $\mu = e^{\mathbf{x}\beta+n\gamma}$, $v = \dfrac{n}{\phi}$, and $z = \dfrac{\Gamma(n/\phi + 1/p)}{\Gamma(n/\phi)}$.

For the random effect, when $\theta^C \sim \mathcal{GIG}(k+1, w, p)$, the density of $\theta^C$ is given as

$$\pi_C(\theta) = \frac{p}{\Gamma(k+1)}\left(\frac{w}{\theta}\right)^{pk+p}\exp\left(-\frac{w^p}{\theta^p}\right)\frac{1}{\theta}$$

where $w = \dfrac{\Gamma(k+1)}{\Gamma(k+1-1/p)}$.

By integrating out the random effect as shown in Jeong and Valdez (2020a), we can show that $\overline{C}|n \sim \mathcal{GB}2\left(k+1, \mu\dfrac{w}{z}, \dfrac{n}{\phi}, p\right)$ as follows:

$$
\begin{aligned}
f(\overline{c}|n) &= \int_0^\infty f(\overline{c}|n,\theta)\pi_C(\theta)d\theta \\
&= \int_0^\infty f(\overline{c}|n,\theta)\frac{p}{\Gamma(k+1)}\left(\frac{w}{\theta}\right)^{pk+p}\exp\left(-\frac{w^p}{\theta^p}\right)\frac{1}{\theta}d\theta \\
&= \frac{p^2\overline{c}^{pv-1}(\frac{z}{\mu})^{pv}w^{pk+p}}{\Gamma(v)\Gamma(k+1)}\int_0^\infty \theta^{-pk-pv-p-1}e^{(-\frac{w^p+(\overline{c}z/\mu)^p}{\theta^p})}d\theta \\
&= \frac{p^2\overline{c}^{pv-1}(\frac{z}{\mu})^{pv}w^{pk+p}}{\Gamma(v)\Gamma(k+1)}|\frac{1}{p}\int_0^\infty x^{-k-v-2}e^{(-\frac{w^p+(\overline{c}z/\mu)^p}{x})}dx| \\
&= \frac{|p|\overline{c}^{pv-1}(z/\mu)^{pv}w^{pk+p}}{\Gamma(v)\Gamma(k+1)}\frac{\Gamma(v+k+1)}{(w^p+(\overline{c}z/\mu)^p)^{v+k+1}} \\
&= |p|\frac{\Gamma(v+k+1)}{\Gamma(v)\Gamma(k+1)}\frac{\overline{c}^{-1}(\overline{c}z/\mu)^{pv}w^{pk+p}}{(w^p+(\overline{c}z/\mu)^p)^{v+k+1}} \\
&= |p|\frac{\Gamma(v+k+1)}{\Gamma(v)\Gamma(k+1)}\frac{\overline{c}^{pv-1}(w\mu/z)^{pk+p}}{(\overline{c}^p+(w\mu/z)^p)^{v+k+1}} \\
&= |p|\frac{\Gamma(\frac{n}{\phi}+k+1)}{\Gamma(\frac{n}{\phi})\Gamma(k+1)}\frac{\overline{c}^{pn/\phi-1}\left(\frac{\Gamma(k+1)\Gamma(\frac{n}{\phi})}{\Gamma(k+1-\frac{1}{p})\Gamma(\frac{n}{\phi}+\frac{1}{p})}\mu\right)^{pk+p}}{\left(\overline{c}^p+(\frac{\Gamma(k+1)\Gamma(\frac{n}{\phi})}{\Gamma(k+1-\frac{1}{p})\Gamma(\frac{n}{\phi}+\frac{1}{p})}\mu)^p\right)^{n/\phi+k+1}},
\end{aligned}
$$

where in the third line, we have $x := \theta^p$ and in the fourth line, we have $\frac{dx}{d\theta} = p\theta^{p-1}$. Note that if $Y \sim \mathcal{GB}2(a,\xi,\tau,p)$, then the density function is given as follows:

$$
f(y|a,\xi,\tau,p) = \frac{\Gamma(a+\tau)}{\Gamma(a)\Gamma(\tau)}|p|\frac{\xi^{ap}y^{\tau p-1}}{(y^p+\xi^p)^{a+\tau}}.
$$

We know that $\theta^C|\mathbf{n}_T,\overline{\mathbf{c}}_T \sim \mathcal{GIG}(k_T+1, w^*_{T,p}, p)$, where $z_t = \dfrac{\Gamma(n_t/\phi+1/p)}{\Gamma(n_t/\phi)}$, $w = \dfrac{\Gamma(k+1)}{\Gamma(k+1-1/p)}$, $w^*_{T,p} = [w^p + \sum_{t=1}^T(z_t\overline{c}_t\mu_t^{-1})^p]^{1/p}$, $k_T = k + \sum_{t=1}^T n_t/\phi$ from Appendix A.

Thus, based on the derivation of the marginal GB2 density, it can be shown that $\overline{C}_{T+1}|\mathbf{n}_{T+1},\overline{\mathbf{c}}_T \sim \mathcal{GB}2\left(k_T+1, \mu_{T+1}\dfrac{w^*_{T,p}}{z_{T+1}}, \dfrac{n_{T+1}}{\phi}, p\right)$ and

$$
\begin{aligned}
\mathbb{E}\left[\overline{C}_{T+1}|\mathbf{n}_{T+1},\overline{\mathbf{c}}_T\right] &= w^*_{T,p}\frac{\Gamma(k_T+1-1/p)}{\Gamma(k_T+1)}\mu_{T+1} \\
&= \left(\sum_{t=1}^T\left(\frac{S_t}{\mu_t}\frac{\Gamma(n_t/\phi+1/p)}{\phi\Gamma(n_t/\phi+1)}\right)^p + w^p\right)^{1/p}\frac{\Gamma(k_T+1-1/p)}{\Gamma(k_T+1)}\mu_{T+1}.
\end{aligned}
$$

# Appendix C. Proof of Lemma 4.4

First, we plug $F_t(y_t)$ in $u_t$ as $F_t(y_t) = B_{\psi,k}\left(\dfrac{y_t^p}{y_t^p + c^p}\right)$ so that

$$q_t = B_{\psi,k}^{-1}(F_t(y_t))/[1 - B_{\psi,k}^{-1}(F_t(y_t))] = (y_t/c)^p.$$

We now have

$$c_T(\mathbf{u}_T)\Big|_{\mathbf{u}=F(\mathbf{y})} = \frac{\Gamma(k)^{T-1}\Gamma(\psi T + k)}{\Gamma(\psi + k)^T} \times \frac{\prod_{t=1}^{T}(1 + (y_t/c)^p)^{\psi+k}}{(1 + \sum_{t=1}^{T}(y_t/c)^p)^{\psi T + k}}.$$

It follows that we can simplify $\left.\dfrac{c_{T+1}(\mathbf{u}_{T+1})}{c_T(\mathbf{u}_T)}\right|_{\mathbf{u}=F(\mathbf{y})}$ as follows:

$$\begin{aligned}
\left.\frac{c_{T+1}(\mathbf{u}_{T+1})}{c_T(\mathbf{u}_T)}\right|_{\mathbf{u}=F(\mathbf{y})} &= \frac{\Gamma(k)^T\Gamma(\psi(T+1) + k)}{\Gamma(\psi + k)^{T+1}} \times \frac{\prod_{t=1}^{T+1}(1 + (y_t/c)^p)^{\psi+k}}{(1 + \sum_{t=1}^{T+1}(y_t/c)^p)^{\psi(T+1)+k}} \\
&\div \frac{\Gamma(k)^{T-1}\Gamma(\psi T + k)}{\prod_{t=1}^{T}\Gamma(\psi + k)} \times \frac{\prod_{t=1}^{T}(1 + (y_t/c)^p)^{\psi+k}}{(1 + \sum_{t=1}^{T}(y_t/c)^p)^{\psi T + k}} \\
&= \frac{\Gamma(k)\Gamma(\psi + k_T)}{\Gamma(\psi + k)\Gamma(k_T)} \times \frac{(1 + (y_{T+1}/c)^p)^{\psi+k}}{(1 + \sum_{t=1}^{T+1}(y_t/c)^p)^{\psi+k_T}} \times (1 + \sum_{t=1}^{T}(y_t/c)^p)^{k_T} \\
&= \frac{\Gamma(k)\Gamma(\psi + k_T)}{\Gamma(\psi + k)\Gamma(k_T)} \times \frac{(c^p + y_{T+1}^p)^{\psi+k}}{(c^p + \sum_{t=1}^{T+1} y_t^p)^{\psi+k_T}} \times \frac{(c^p + \sum_{t=1}^{T} y_t^p)^{k_T}}{c^{pk}} \\
&= \frac{\Gamma(k)\Gamma(\psi + k_T)}{\Gamma(\psi + k)\Gamma(k_T)} \times \frac{(c^p + y_{T+1}^p)^{\psi+k}}{({c_{T,p}^*}^p + y_{T+1}^p)^{\psi+k_T}} \times \frac{{c_{T,p}^*}^{pk_T}}{c^{pk}}
\end{aligned}$$

where $c_{T,p}^* = \left(c^p + \sum_{t=1}^{T} y_t^p\right)^{1/p}$ and $k_T = k + \psi T$.

Note that $f_{T+1}(y_{T+1})$ is given as

$$\begin{aligned}
f_{T+1}(y_{T+1}) &= \int_0^\infty f_{Y_{T+1}}(y_{T+1}|\theta)p(\theta)d\theta \\
&= \frac{p}{y_{T+1}\,B(\psi,k)}\frac{c^{pk}\,y_{T+1}^{p\psi}}{\left(c^p+y_{T+1}^p\right)^{\psi+k}} = \frac{p}{y_{T+1}}\frac{\Gamma(\psi+k)}{\Gamma(\psi)\Gamma(k)}\frac{c^{pk}\,y_{T+1}^{p\psi}}{\left(c^p+y_{T+1}^p\right)^{\psi+k}}.
\end{aligned}$$

Thus, we have

$$\begin{aligned}
f_{T+1}(y_{T+1})\frac{c_{T+1}(\mathbf{u}_{T+1})}{c_T(\mathbf{u}_T)}\bigg|_{\mathbf{u}=F(\mathbf{y})} &= \frac{p}{y_{T+1}}\frac{\Gamma(\psi+k)}{\Gamma(\psi)\Gamma(k)}\frac{c^{pk}\,y_{T+1}^{p\psi}}{\left(c^p+y_{T+1}^p\right)^{\psi+k}} \\
&\quad\times\frac{\Gamma(k)\Gamma(\psi+k_T)}{\Gamma(\psi+k)\Gamma(k_T)}\times\frac{(c^p+y_{T+1}^p)^{\psi+k}}{({c_{T,p}^*}^p+y_{T+1}^p)^{\psi+k_T}}\times\frac{{c_{T,p}^*}^{pk_T}}{c^{pk}} \\
&= \frac{p}{y_{T+1}}\frac{\Gamma(\psi+k_T)}{\Gamma(\psi)\Gamma(k_T)}\times\frac{{c_{T,p}^*}^{pk_T}\,y_{T+1}^{p\psi}}{({c_{T,p}^*}^p+y_{T+1}^p)^{\psi+k_T}} \\
&= \frac{p}{y_{T+1}\,B(\psi,k_T)}\times\frac{{c_{T,p}^*}^{pk_T}\,y_{T+1}^{p\psi}}{({c_{T,p}^*}^p+y_{T+1}^p)^{\psi+k_T}}.
\end{aligned}$$

# Appendix D. Proof of Lemmas 5.1 and 5.2

**Lemma 6.1.** *If* $\left(R, \boldsymbol{Z}_{(\boldsymbol{\tau})}\right)$ *follows* (5.12)*, then*

$$\Phi^{-1}(F_t(N_t))|R, y_{t,1}, \ldots, y_{t,n_t} \sim \mathrm{N}(\mu_t\left(n_t\right), (\sigma\left(n_t\right))^2) \tag{D.1}$$

*where*

$$\mu_t\left(n_t\right) = \left(\theta_1, \rho_1, \cdots, \rho_1\right)\left(\boldsymbol{\Sigma}_{tt}^{(\boldsymbol{\rho}^*)}\right)^{-1}\left(r, \Phi^{-1}\left(G(y_{t,1})\right), \cdots, \Phi^{-1}\left(G(y_{t,n_t})\right)\right)^{\mathrm{T}}$$

*and*

$$\left(\sigma\left(n_t\right)\right)^2 = 1 - \left(\theta_1, \rho_1, \cdots, \rho_1\right)\left(\boldsymbol{\Sigma}_{tt}^{(\boldsymbol{\rho}^*)}\right)^{-1}\left(\theta_1, \rho_1, \cdots, \rho_1\right)^{\mathrm{T}}$$

*with* $\rho_1^* = \theta_2$*,* $\rho_2^* = \rho_2$*. Furthermore, we have*

$$\mathbb{P}(N_t \leq n|r, y_{t,1}, \cdots, y_{t,n_t}) = \Phi\left(\frac{\Phi^{-1}(F_t(n)) - \mu_t\left(n_t\right)}{\sigma\left(n_t\right)}\right). \tag{D.2}$$

*Proof.* Since $\left(R, \boldsymbol{Z}_{(\boldsymbol{\tau})}\right)$ is properly nested in terms of $\boldsymbol{n}$, we can consider the copula structure of $(R, \boldsymbol{Z}_t)$ as a special case of multi-year random effects model. Therefore, one can show that $(N_t, R, Y_{t,1}, \ldots, Y_{t,n_t})$ has the following structure:

$$(N_t, R, Y_{t,1}, \ldots, Y_{t,n_t}) \sim C_{\Sigma_t}\left(F_t, \Phi, G_{t,1}, \cdots, G_{t,n_t}\right)$$

where

$$[\Sigma_t]_{\ell m} = \begin{cases} 1, & \ell = m; \\ \theta_1, & \ell + m = 3; \\ \rho_1, & 1 = \ell, m > 2 \text{ or } \ell > 2, m = 1; \\ \theta_2, & 2 = \ell < m \text{ or } \ell > m = 2; \\ \rho_2, & \text{elsewhere}; \end{cases} = \left(\begin{array}{c|c|cccc} 1 & \theta_1 & \rho_1 & \cdots & \cdots & \rho_1 \\ \hline \theta_1 & 1 & \theta_2 & \cdots & \cdots & \theta_2 \\ \hline \rho_1 & \theta_2 & 1 & \rho_2 & \cdots & \rho_2 \\ \vdots & \vdots & \rho_2 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & 1 & \rho_2 \\ \rho_1 & \theta_2 & \rho_2 & \cdots & \rho_2 & 1 \end{array}\right).$$

By definition of Gaussian copula, it is easy to see that

$$\left(\Phi^{-1}(F_t(N_t)), R, \Phi^{-1}(G_{t,1}(Y_{t,1})), \cdots, \Phi^{-1}(G_{t,n_t}(Y_{t,n_t}))\right) \sim \mathrm{N}(\mathbf{0}_{n_t+2}, \boldsymbol{\Sigma}_t).$$

From the property of multivariate normal distribution, if $(Z_1, Z_2) \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} \text{ and } \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{1,1} & \boldsymbol{\Sigma}_{1,2} \\ \boldsymbol{\Sigma}_{2,1} & \boldsymbol{\Sigma}_{2,2} \end{pmatrix},$$

then

$$Z_1 | Z_2 = z_2 \sim \mathrm{N}\left(\boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{1,2}\boldsymbol{\Sigma}_{2,2}^{-1}(z_2 - \boldsymbol{\mu}_2),\ \boldsymbol{\Sigma}_{1,1} - \boldsymbol{\Sigma}_{1,2}\boldsymbol{\Sigma}_{2,2}^{-1}\boldsymbol{\Sigma}_{2,1}\right).$$

Therefore, $\Phi^{-1}(F_t(N_t))|R, Y_{t,1}, \ldots, Y_{t,n_t}$ follows a univariate normal distribution so that

$$\Phi^{-1}(F_t(N_t))|R = r, Y_{t,1} = y_{t,1}, \ldots, Y_{t,n_t} = y_{t,n_t} \sim \mathrm{N}(\mu_t(n_t), (\sigma(n_t))^2)$$

where

$$\mu_t(n_t) = (\theta_1, \rho_1, \cdots, \rho_1)\left(\boldsymbol{\Sigma}_{tt}^{(\boldsymbol{\rho}^*)}\right)^{-1}\left(r, \Phi^{-1}(G(y_{t,1})), \cdots, \Phi^{-1}(G(y_{t,n_t}))\right)^{\mathrm{T}}$$

and

$$\left(\sigma\left(n_t\right)\right)^2 = 1 - \left(\theta_1, \rho_1, \cdots, \rho_1\right) \left(\mathbf{\Sigma}_{tt}^{(\boldsymbol{\rho}^*)}\right)^{-1} \left(\theta_1, \rho_1, \cdots, \rho_1\right)^{\mathrm{T}}$$

with $\rho_1^* = \theta_2,\ \rho_2^* = \rho_2$ because

$$\left(\mathbf{\Sigma}_{tt}^{(\boldsymbol{\rho}_*)}\right)\Bigg|_{\rho_1^*=\theta_2,\,\rho_2^*=\rho_2} = \left(\begin{array}{c:cccc} 1 & \theta_2 & \cdots & \cdots & \theta_2 \\ \hdashline \theta_2 & 1 & \rho_2 & \cdots & \rho_2 \\ \vdots & \rho_2 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & 1 & \rho_2 \\ \theta_2 & \rho_2 & \cdots & \rho_2 & 1 \end{array}\right).$$

□

**Lemma 6.2.** *Consider the settings in* (5.12)*. Then, the density function of* $(R, \boldsymbol{Z}_t)$ *conditional on* $R = r$ *is given by*

$$g^*_{t|R}\left(y_{t,1}, \cdots, y_{t,n_t}|r\right) = \phi_{\boldsymbol{\mu}^*,\mathbf{\Sigma}^*}\left(\Phi^{-1}\left(G\left(y_{t,1}\right)\right), \cdots, \Phi^{-1}\left(G\left(y_{t,n_t}\right)\right)\right) \prod_{j=1}^{n_t} \frac{g(y_{t,j})}{\phi\left(\Phi^{-1}\left(G(y_{t,j})\right)\right)}, \tag{D.3}$$

*where*

$$\mu^* = r\,\theta_2 \mathbf{1}_{n_t} \quad \textit{and} \quad \mathbf{\Sigma}^* = (1-\rho_2)\boldsymbol{I}_{n_t} + (\rho_2 - \theta_2^2)\boldsymbol{J}_{n_t \times n_t}.$$

*Proof.* Note that

$$\left(R, Y_{t,1}, \ldots, Y_{t,n_t}\right) \sim C_{tt}^{(\boldsymbol{\rho}^*)}\left(\Phi, G_{t,1}, \cdots, G_{t,n_t}\right)$$

with $\rho_1^* = \theta_2,\ \rho_2^* = \rho_2$, where $C_{tt}^{(\boldsymbol{\rho}^*)}$ is a Gaussian copula with correlation matrix $\mathbf{\Sigma}_{tt}^{(\boldsymbol{\rho}^*)}$

Thus, one can show that $\Phi^{-1}(G_{t,1}(y_{t,1})), \cdots, \Phi^{-1}(G_{t,n_t}(y_{t,n_t}))|R$ follows a multivariate normal distribution, with mean vector given as

$$\left(\theta_2, \ldots, \theta_2\right)^{\mathrm{T}} \cdot {I_1}^{-1} \cdot r = r\theta_2 \mathbf{1}_{n_t},$$

and covariance matrix given as

$$(1-\rho_2)\boldsymbol{I}_{n_t} + \rho_2 \boldsymbol{J}_{n_t\times n_t} - (\theta_2,\ldots,\theta_2)\cdot(\theta_2,\ldots,\theta_2)^{\mathrm{T}} = (1-\rho_2)\boldsymbol{I}_{n_t} + (\rho_2-\theta_2^2)\boldsymbol{J}_{n_t\times n_t}.$$

Therefore, we have

$$G^*_{t|R}\left(y_{t,1},\cdots,y_{t,n_t}|r\right) = \Phi_{\mu^*,\boldsymbol{\Sigma}^*}\left(\Phi^{-1}\left(G\left(y_{t,1}\right)\right),\cdots,\Phi^{-1}\left(G\left(y_{t,n_t}\right)\right)\right).$$

Finally, one can obtain the following formula by differentiating $G^*_{t|R}\left(y_{t,1},\cdots,y_{t,n_t}|r\right)$ with respect to $y_{t,1},\cdots,y_{t,n_t}$ and then applying the chain rule:

$$g^*_{t|R}\left(y_{t,1},\cdots,y_{t,n_t}|r\right) = \phi_{\mu^*,\boldsymbol{\Sigma}^*}\left(\Phi^{-1}\left(G\left(y_{t,1}\right)\right),\cdots,\Phi^{-1}\left(G\left(y_{t,n_t}\right)\right)\right)\prod_{j=1}^{n_t}\frac{g(y_{t,j})}{\phi\left(\Phi^{-1}\left(G(y_{t,j})\right)\right)}.$$

□

# Appendix E. Multi-year microlevel collective risk model with $t$ copulas

Let $F$ and $F_t$ be non-negative integer-valued distribution functions with the respective probability mass functions $f_0$ and $f_t$ for $t \in \mathbb{N}$, respectively. Let $G$ and $G_{t,j}$ be non-negative real-valued distribution functions with respective probability densities $g$ and $g_{t,j}$ for $t, j \in \mathbb{N}$.

For $\boldsymbol{\rho}$ defined in (5.5) satisfying (5.7), we define the joint distribution of $\boldsymbol{Z}_{(\boldsymbol{\tau})}$ as

$$\boldsymbol{Z}_{(\boldsymbol{\tau})} \sim H = C^{\nu,(\boldsymbol{\rho})}_{(\boldsymbol{n})} \left(F_1, G_{1,1}, \cdots, G_{1,n_1}, F_2, G_{2,1}, \cdots, G_{2,n_2}, \cdots, F_\tau, G_{\tau,1}, \cdots, G_{\tau,n_\tau}\right), \quad \text{(E.1)}$$

repeatedly for any $\boldsymbol{n} = (n_1, \cdots, n_\tau) \in (\mathbb{N}_0)^\tau$, where $C^{\nu;\boldsymbol{\rho}}_{(\boldsymbol{n})}$ is a $t$ copula with scale matrix $\boldsymbol{\Sigma}^{(\boldsymbol{\rho})}_{(\boldsymbol{n})}$ and degree of freedom $\nu$. While it is not necessary but for simplicity, we assume $F_t = F$ and $G_{tj} = G$ for any $t, j \in \mathbb{N}$.

Let $\Phi_\nu$ and $\phi_\nu$ be the cumulative distribution function and density function of univariate standard t-distribution with mean 0, standard variation 1, and degree of freedom $\nu$, respectively. Similar to (5.10), the density function for (E.1) requires multiple integration of the $t$ copula depending on the length of the observation years. To detour from such difficulty, we have the following result. Note that Corollary 5.1 is a special case of Corollary 6.1 when $\nu = \infty$.

**Corollary 6.1.** *Consider the specification of $\boldsymbol{Z}_{(\boldsymbol{\tau})}$ in (E.1). Then, we have*

$$\begin{aligned} &h(\boldsymbol{z}_{(\boldsymbol{\tau})}) \\ &= \int\int \prod_{t=1}^{\tau} \Bigg[ g^*_{t|R,W}\left(y_{t,1}, \cdots, y_{t,n_t} | r, w\right) \\ &\qquad \left( \Phi\left( \frac{\Phi^{-1}(F(n_t)) - \mu_t\left(n_t\right)}{\sigma\left(n_t\right)} \right) - \Phi\left( \frac{\Phi^{-1}(F(n_t - 1)) - \mu_t\left(n_t\right)}{\sigma\left(n_t\right)} \right) \right) \Bigg] \phi(r) \nu f_\nu(w\,\nu)\, \mathrm{d}r \mathrm{d}w \end{aligned} \quad \text{(E.2)}$$

*where $f_\nu$ is a density function of chi-squared distribution with $\nu$ degrees of freedom, and*

$$\mu_t\left(n_t\right) = \left(\theta_1, \rho_1, \cdots, \rho_1\right) \left(\boldsymbol{\Sigma}_{tt}^{(\boldsymbol{\rho}^*)}\right)^{-1} \left(r, \Phi^{-1}\left(G(y_{t,1})\right), \cdots, \Phi^{-1}\left(G(y_{t,n_t})\right)\right)^{\mathrm{T}}$$

*and*

$$\left(\sigma\left(n_t\right)\right)^2 = \frac{1 - \left(\theta_1, \rho_1, \cdots, \rho_1\right) \left(\boldsymbol{\Sigma}_{tt}^{(\boldsymbol{\rho}^*)}\right)^{-1} \left(\theta_1, \rho_1, \cdots, \rho_1\right)^{\mathrm{T}}}{w}$$

*with $\rho_1^* = \theta_2$ and $\rho_2^* = \rho_2$. Here, $g^*_{t|R,W}\left(\cdot|r\right)$ is the density function of $\boldsymbol{Y}_t$ conditional on $R = r$ and $W = w$, and given by*

$$\begin{aligned} & g^*_{t|R,W}\left(y_{t,1}, \cdots, y_{t,n_t}|r, w\right) \\ & = \phi_{\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*}\left(\Phi^{-1}\left(G\left(y_{t,1}\right)\right), \cdots, \Phi^{-1}\left(G\left(y_{t,n_t}\right)\right)\right) \prod_{j=1}^{n_t} \frac{g(y_{t,j})}{\phi\left(\Phi^{-1}\left(G(y_{t,j})\right)\right)} \end{aligned}$$

*where*

$$\boldsymbol{\mu}^* = r\,\theta_2 \mathbf{1}_{n_t} \quad \text{and} \quad \boldsymbol{\Sigma}^* = \frac{(1-\rho_2)\boldsymbol{I}_{n_t} + (\rho_2 - \theta_2^2)\boldsymbol{J}_{n_t \times n_t}}{w}.$$

*Proof of Corollary 6.1.* Knowing that multivariate $t$-distribution with the degree of freedom $\nu$ can be represented as a multivariate normal distribution conditional on the latent variable $W$ whose density function at $W = w$ is given by $\nu f_\nu(w\,\nu)$, the proof follows immediately from Corollary 5.1. □

# Appendix F. Mathematical Justification of Model 5.1

As briefly discussed in Remark 5.1, Model 5.1 is not casual in the sense that the length or dimension of the observation varies depending on the value of the observation. One solution to detour the difficulty from the varying dimension of the observation is that we may assume the infinite number of severities regardless of the value of the frequency $n_t$. Specifically, we define

$$\boldsymbol{Z}_t(\boldsymbol{k}_t) := (N_1, \boldsymbol{Y}_1(k_1), \cdots, N_t, \boldsymbol{Y}_t(k_t))$$

for any $\boldsymbol{k}_t := (k_1, \cdots, k_t) \in \mathbb{N}_0^t$. Then, $\boldsymbol{y}_t = \boldsymbol{y}_t(n_t)$ can be understood as the observation where we only observe first $n_t$ severities among the infinite number of severities. Then, Model 5.1 can be reformulated as follows.

**Model 6.1** (Revision of Model 5.1)**.** *We repeatedly define the following random effect model for all possible values of*

$$k_t \in \mathbb{N}_0, \quad t = 1, \cdots \tag{F.1}$$

*where the joint distribution between observations and the random effect model is presented with the copula model with parts i and iv are the same as in Model 5.1.*

*ii. Conditional on $R = r$, we have that $\boldsymbol{Z}_t(k_t)$ for $t = 1, \cdots$ are independent observations whose distribution function is given by*

$$H_t^{\#}\left(\boldsymbol{z}_t(k_t)|r\right) := C_{(\theta_3,\theta_4)}\left(F_t(n_t|r), G_t(y_{t,1}|r), \ldots, G_t(y_{t,k_t}|r)\right).$$

*As a result, we have the following distribution function of $\boldsymbol{Z}_{(\tau)}(\boldsymbol{k}_\tau)$*

$$H^{\#}(\boldsymbol{z}_{(\tau)}(\boldsymbol{k}_\tau)) := \int \prod_{t=1}^{\tau} H_t^{\#}\left(\boldsymbol{z}_t(k_t)|r\right)\pi(r)\mathrm{d}r.$$

*iii. The parameters $\theta_3$ and $\theta_4$ of the copula $C_{(\theta_3,\theta_4)}$ controls the independence between the frequency and severities and independence among individual severities, respectively, within a year so that we have*

$$h_t^{\#}(\boldsymbol{z}_t(k_t)|r) = f_t(n_t|r) g_t^{[\text{joint}]\#}(\boldsymbol{y}_t(k_t)|r) \quad \textit{if and only if} \quad \theta_3 = 0,$$

*where*

$$g_t^{[\text{joint}]\#}(\boldsymbol{y}_t(k_t)|r) = \prod_{j=1}^{k_t} g_t(y_{t,j}|r) \quad \textit{if and only if} \quad \theta_4 = 0,$$

*where $g_t^{[\text{joint}]\#}$ means joint density function of $\boldsymbol{Y}_t(k_t)$.*

*v.* $\boldsymbol{y}_t(k_t) \perp R$ *for all* $t = 1, \ldots, \tau$ *if and only if* $\theta_2 = 0$.

*vi. (Inheritance Property) Consider two distribution functions*

$$H^{[1]}\left(\boldsymbol{z}_t(k_t)|r\right) := H_t^{\#}\left(\boldsymbol{z}_t(k_t)|r\right) \quad \textit{and} \quad H^{[2]}\left(\boldsymbol{z}_t(k_t^*)|r\right) := H_t^{\#}\left(\boldsymbol{z}_t(k_t^*)|r\right)$$

*for* $k_t \leq k_t^*$. *Then, we have the following inheritance property*

$$H^{[1]}\left(n_t, y_{t,1}, \cdots, y_{t,k_t}|r\right) = \lim_{y_{t,k_t+1}\to\infty, \cdots, y_{t,k_t^*}\to\infty} H^{[2]}\left(n_t, y_{t,1}, \cdots, y_{t,k_t}, y_{t,k_t+1}, \cdots, y_{t,k_t^*}|r\right) \tag{F.2}$$

*for any* $\boldsymbol{z}_t(k_t) = (n_t, y_{t,1}, \cdots, y_{t,k_t})$ *and* $r$.

Note that part v in Model 6.1 is necessary for the well-definedness of the model since the model is repeatedly defined for multiple times for (F.1). One immediate result from Model 6.1 is that its density function at $\boldsymbol{Z}_t(k_t) = \boldsymbol{z}_t(k_t)$

$$h_t^{\#}(\boldsymbol{z}_t(k_t)) = \int h_t^{\#}(n_t, \boldsymbol{y}_t(k_t)|r)\pi(r)\mathrm{d}r$$

is well-defined under the classical multivariate analysis with the following relation with the corresponding joint distribution function

$$H_t^{\#}(\boldsymbol{z}_t(k_t)) = \sum_{x_0=0}^{n_t} \int_{-\infty}^{y_{t,1}} \cdots \int_{-\infty}^{y_{t,k_t}} h_t^{\#}(x_0, x_1, \cdots, x_{k_t})\mathrm{d}x_1 \cdots \mathrm{d}x_{k_t}.$$

Furthermore, importantly, we observe that the density function $h_t^{\#}$ at $\boldsymbol{Z}_t(k_t) = (n_t, y_{t,1}, \cdots, y_{t,k_t})$ in Model 6.1 coincides with the density function $h_t$ in Model 5.1 at $\boldsymbol{Z}_t = (n_t, y_{t,1}, \cdots, y_{t,n_t})$ if $k_t = n_t$. Hence, as long as inheritance property in part vi of Model 6.1 holds, we can see that the density function and the corresponding distribution function in Model 5.1 is well-defined having Model 6.1 as background model. For the simplicity of the presentation, this paper only present Model 5.1 without specifying the background model in Model 6.1.

# References


Antonio, K. and Beirlant, J. (2007). Actuarial statistics with generalized linear mixed models. *Insurance: Mathematics and Economics*, 40(1):58 – 76.

Antonio, K. and Valdez, E. A. (2012). Statistical concepts of a priori and a posteriori risk classification in insurance. *Advances in Statistical Analysis*, 96(2):187–224.

Bailey, A. L. (1950). Credibility procedures: Laplace's generalization of bayes' rule and the combination of collateral knowledge with observed data. In *Proceedings of the Casualty Actuarial Society*, volume 37, pages 7–23.

Baumgartner, C., Gruber, L. F., and Czado, C. (2015). Bayesian total loss estimation using shared random effects. *Insurance: Mathematics and Economics*, 62:194–201.

Berger, J., Berliner, L. M., et al. (1986). Robust bayes and empirical bayes analysis with $\epsilon$-contaminated priors. *The Annals of Statistics*, 14(2):461–486.

Berger, J. O. (1985). *Statistical decision theory and Bayesian analysis*. Springer Science & Business Media.

Berger, J. O., Moreno, E., Pericchi, L. R., Bayarri, M. J., Bernardo, J. M., Cano, J. A., De la Horra, J., Martín, J., Ríos-Insúa, D., Betrò, B., et al. (1994). An overview of robust bayesian analysis. Technical Report 1.

Bolancé, C., Guillén, M., and Pinquet, J. (2003). Time-varying credibility for frequency risk models: estimation and tests for autoregressive specifications on the random effects. *Insurance: Mathematics and Economics*, 33(2):273–282.

Boratyńska, A. (2017). Robust bayesian estimation and prediction of reserves in exponential model with quadratic variance function. *Insurance: Mathematics and Economics*, 76:135–140.

Boucher, J.-P., Denuit, M., and Guillén, M. (2008). Models of insurance claim counts with time dependence based on generalization of poisson and negative binomial distributions. *Variance*, 2(1):135–162.

Boudreault, M., Cossette, H., Landriault, D., and Marceau, E. (2006). On a risk model with dependence between interclaim arrivals and claim sizes. *Scandinavian Actuarial Journal*, 2006(5):265–285.

Bregman, L. M. (1967). The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR computational mathematics and mathematical physics*, 7(3):200–217.

Bühlmann, H. (1967). Experience rating and credibility. *ASTIN Bulletin: The Journal of the International Actuarial Association (IAA)*, 4(3):199–207.

Bühlmann, H. and Gisler, A. (2006). *A course in credibility theory and its applications*. Springer Science & Business Media.

Cossette, H., Marceau, E., and Mtalai, I. (2019). Collective risk models with dependence. *Insurance: Mathematics and Economics*, 87:153–168.

Czado, C., Kastenmeier, R., Brechmann, E. C., and Min, A. (2012). A mixed copula model for insurance claims and claim sizes. *Scandinavian Actuarial Journal*, 2012(4):278–305.

Denuit, M., Maréchal, X., Pitrebois, S., and Walhin, J.-F. (2007). *Actuarial modelling of claim counts: Risk classification, credibility and bonus-malus systems*. John Wiley & Sons.

Ding, J. and Wang, J.-L. (2008). Modeling longitudinal data with nonparametric multiplicative random effects jointly with survival data. *Biometrics*, 64(2):546–556.

Dionne, G. and Vanasse, C. (1989). A generalization of automobile insurance rating models: the negative binomial distribution with a regression component. *ASTIN Bulletin: The Journal of the IAA*, 19(2):199–212.

Doss, D. (1979). Definition and characterization of multivariate negative binomial distribution. *Journal of Multivariate Analysis*, 9(3):460–464.

Dubitzky, W., Wolkenhauer, O., Yokota, H., and Cho, K.-H. (2013). *Encyclopedia of systems biology*. Springer.

Eguchi, S. and Kano, Y. (2001). Robustifying maximum likelihood estimation. Technical report, Institute of Statistical Mathematics.

Eichenauer, J., Lehn, J., and Rettig, S. (1988). A gamma-minimax result in credibility theory. *Insurance: Mathematics and Economics*, 7(1):49–57.

Frangos, N. E. and Vrontos, S. D. (2001). Design of optimal bonus-malus systems with a frequency and a severity component on an individual basis in automobile insurance. *ASTIN Bulletin: The Journal of the IAA*, 31(1):1–22.

Frees, E. W., Derrig, R. A., and Meyers, G. (2014a). *Predictive Modeling Applications in Actuarial Science, Volume 1: Predictive Modeling Techniques*. Cambridge University Press: Cambridge, U.K.

Frees, E. W., Derrig, R. A., and Meyers, G. (2016a). *Predictive Modeling Applications in Actuarial Science, Volume 2: Case Studies in Insurance*. Cambridge University Press, Cambridge, U.K.

Frees, E. W., Gao, J., and Rosenberg, M. (2011a). Predicting the frequency and amount of health care expenditures. *North American Actuarial Journal*, 15(3):377 – 392.

Frees, E. W. and Kim, J.-S. (2006). Multilevel model prediction. *Psychometrika*, 71(1):79–104.

Frees, E. W., Lee, G., and Yang, L. (2016b). Multivariate frequency-severity regression models in insurance. *Risks*, 4(1):4.

Frees, E. W., Meyers, G., and Cummings, A. D. (2011b). Summarizing insurance scores using a gini index. *Journal of the American Statistical Association*, 106(495):1085–1098.

Frees, E. W., Meyers, G., and Cummings, A. D. (2014b). Insurance ratemaking and a gini index. *Journal of Risk and Insurance*, 81(2):335–366.

Frees, E. W., Meyers, G., and Cummings, A. D. (2014c). Insurance ratemaking and a gini index. *Journal of Risk and Insurance*, 81(2):335–366.

Frees, E. W., Shi, P., and Valdez, E. A. (2009). Actuarial applications of a hierarchical insurance claims model. *ASTIN Bulletin: The Journal of the IAA*, 39(1):165–197.

Frees, E. W. and Valdez, E. A. (1998). Understanding relationships using copulas. *North American Actuarial Journal*, 2(1):1–25.

Frees, E. W. and Valdez, E. A. (2008). Hierarchical insurance claims modeling. *Journal of the American Statistical Association*, 103(484):1457 – 1469.

Frees, E. W., Young, V. R., and Luo, Y. (1999). A longitudinal data analysis interpretation of credibility models. *Insurance: Mathematics and Economics*, 24(3):229–247.

Garrido, J., Genest, C., and Schulz, J. (2016). Generalized linear models for dependent frequency and severity of insurance claims. *Insurance: Mathematics and Economics*, 70:205 – 215.

Gelfand, A. and Dey, D. (1991). On bayesian robustness of contaminated classes of priors. *Statistics & Risk Modeling*, 9(1-2):63–80.

Genest, C. and Nešlehová, J. (2007). A primer on copulas for count data. *ASTIN Bulletin: The Journal of the IAA*, 37(2):475–515.

Genz, A. and Bretz, F. (2009). *Computation of multivariate normal and t probabilities*, volume 195. Springer Science & Business Media.

Goh, G. and Dey, D. K. (2014). Bayesian model diagnostics using functional bregman divergence. *Journal of Multivariate Analysis*, 124:371–383.

Gómez-Déniz, E. (2016). Bivariate credibility bonus–malus premiums distinguishing between two types of claims. *Insurance: Mathematics and Economics*, 70:117–124.

Gómez-Déniz, E., Bermúdez, L., and Morillo, I. (2005). Computing bonus–malus premiums under partial prior information. *British Actuarial Journal*, 11(2):361–374.

Gómez-Déniz, E., Hernández, A., Pérez, J., and Vázquez-Polo, F. (2002a). Measuring sensitivity in a bonus–malus system. *Insurance: Mathematics and Economics*, 31(1):105–113.

Gómez-Déniz, E., Hernández-Bastida, A., and Vázquez-Polo, F. J. (1999). The esscher premium principle in risk theory: a bayesian sensitivity study. *Insurance: Mathematics and Economics*, 25(3):387–395.

Gómez-Déniz, E., Hernandez-Bastida, A., and Vázquez-Polo, F. J. (2002b). Bounds for ratios of posterior expectations: applications in the collective risk model. *Scandinavian Actuarial Journal*, 2002(1):37–44.

Gómez-Déniz, E., Perez-Sanchez, J. M., and Vázquez-Polo, F. J. (2006). On the use of posterior regret $\gamma$-minimax actions to obtain credibility premiums. *Insurance: Mathematics and Economics*, 39(1):115–121.

Gómez-Déniz, E. and Vázquez-Polo, F. J. (2005). Modelling uncertainty in insurance bonus–malus premium principles by using a bayesian robustness approach. *Journal of Applied Statistics*, 32(7):771–784.

Gómez-Déniz, E., Vázquez Polo, F. J., and Bastida, A. H. (2000). Robust bayesian premium principles in actuarial science. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 49(2):241–252.

Gurrin, L. C., Blake, K. V., Evans, S. F., and Newnham, J. P. (2001). Statistical measures of foetal growth using linear mixed models applied to the foetal origins hypothesis. *Statistics in medicine*, 20(22):3391–3409.

Hausman, J. A., Hall, B. H., and Griliches, Z. (1984). Econometric models for count data with an application to the patents-r&d relationship.

Hernández-Bastida, A., Fernández-Sánchez, M., and Gómez-Déniz, E. (2009). The net bayes premium with dependence between the risk profiles. *Insurance: Mathematics and Economics*, 45(2):247–254.

Hougaard, P., Harvald, B., and Holm, N. V. (1992). Measuring the similarities between the lifetimes of adult danish twins born between 1881–1930. *Journal of the American Statistical Association*, 87(417):17–24.

Insua, S. R., Martin, J., Insua, D. R., and Ruggeri, F. (1999). Bayesian forecasting for accident proneness evaluation. *Scandinavian Actuarial Journal*, 1999(2):134–156.

Jaba, E., Robu, I.-B., and Balan, C. B. (2017). Panel data analysis applied in financial performance assessment. *Romanian Statistical Review*, (2).

Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 186(1007):453–461.

Jeong, H. (2020). Testing the presence of random effects via Bregman divergence. *ASTIN Bulletin: The Journal of the IAA*. accepted for publication.

Jeong, H., Ahn, J., Park, S., and Valdez, E. A. (2020). Generalized linear mixed models for dependent compound risk models. *Variance*. accepted for publication.

Jeong, H., Gan, G., and Valdez, E. (2018). Association rules for understanding policyholder lapses. *Risks*, 6(3):69.

Jeong, H. and Valdez, E. A. (2020a). Bayesian credibility premium with GB2 copulas. *Dependence Modeling*. accepted for publication.

Jeong, H. and Valdez, E. A. (2020b). Predictive compound risk models with dependence. *SSRN, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3494575*.

Jewell, W. S. (1974). Credible means are exact bayesian for exponential families. *ASTIN Bulletin: The Journal of the IAA*, 8(1):77–90.

Joe, H. (1997). *Multivariate models and multivariate dependence concepts*. Chapman and Hall/CRC.

Kadhem, S. H. and Nikoloulopoulos, A. K. (2019). Factor copula models for mixed data. *arXiv preprint arXiv:1907.07395*.

Karimnezhad, A. and Parsian, A. (2018). Bayesian and robust bayesian analysis in a general setting. *Communications in Statistics-Theory and Methods*, pages 1–29.

Kim, G.-S., Paik, M. C., and Kim, H. (2017). Causal inference with observational data under cluster-specific non-ignorable assignment mechanism. *Computational Statistics & Data Analysis*, 113:88–99.

Klugman, S. A., Panjer, H. H., and Willmot, G. E. (2012). *Loss models: from data to decisions*, volume 715. John Wiley & Sons.

Krämer, N., Brechmann, E. C., Silvestrini, D., and Czado, C. (2013). Total loss estimation using copula-based regression models. *Insurance: Mathematics and Economics*, 53(3):829–839.

Krupskii, P. and Joe, H. (2013). Factor copula models for multivariate data. *Journal of Multivariate Analysis*, 120:85–101.

Krupskii, P. and Joe, H. (2015). Structured factor copula models: Theory, inference and computation. *Journal of Multivariate Analysis*, 138:53–73.

Landsman, Z. M. and Valdez, E. A. (2003). Tail conditional expectations for elliptical distributions. *North American Actuarial Journal*, 7(4):55–71.

Lee, G. Y. and Shi, P. (2019). A dependent frequency–severity approach to modeling longitudinal insurance claims. *Insurance: Mathematics and Economics*, 87:115–129.

Lee, W., Kim, J., and Ahn, J. Y. (2020). The poisson random effect model for experience ratemaking: Limitations and alternative solutions. *Insurance: Mathematics and Economics*, 91:26–36.

Lee, W., Park, S. C., and Ahn, J. Y. (2018). Investigating dependence between frequency and severity via simple generalized linear models. *Journal of the Korean Statistical Society*.

Lee, Y. and Nelder, J. A. (1996). Hierarchical generalized linear models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(4):619–678.

Lemaire, J. (1998). Bonus-malus systems: the european and asian approach to merit-rating. *North American Actuarial Journal*, 2(1):26–38.

Li, D. X. (2000). On default correlation: a copula function approach. *Journal of Fixed Income*, 9:43–54.

Mayerson, A. L. (1964). A bayesian view of credibility. In *Proceedings of the Casualty Actuarial Society*, volume 51, pages 7–23.

McDonald, J. B. and Bookstaber, R. M. (1991). Option pricing for generalized distributions. *Communications in Statistics - Theory and Methods*, 20(12):4053–4068.

McDonald, J. B. and Butler, R. J. (1990). Regression models for positive random variables. *Journal of Econometrics*, 43:227–251.

Mestiri, S. and Hamdi, M. (2012). Credit risk prediction: A comparative study between logistic regression and logistic regression with random effects. *International Journal of Management Science and Engineering Management*, 7(3):200–204.

Molenberghs, G., Verbeke, G., Demétrio, C., and Vieira, A. (2010). A family of generalized linear models for repeated measures with normal and conjugate random effects. *Statistical Science*, 25(3):325–347.

Morris, C. N. (1983). Parametric empirical bayes inference: theory and applications. *Journal of the American Statistical Association*, 78(381):47–55.

Murray, J. S., Dunson, D. B., Carin, L., and Lucas, J. E. (2013). Bayesian gaussian copula factor models for mixed data. *Journal of the American Statistical Association*, 108(502):656–665.

Nelder, J. and Wedderburn, R. (1972). Generalized linear models. *Journal of the Royal Statistical Society. Series A (General)*, 135(3):370 – 384.

Ng, E. and Cook, R. (2000). A comparison of some random effect models for parameter estimation in recurrent events. *Mathematical and computer modelling*, 32(1-2):11–26.

Nikoloulopoulos, A. K. and Joe, H. (2015). Factor copula models for item response data. *Psychometrika*, 80(1):126–150.

Norberg, R. (1986). Hierarchical credibility: analysis of a random effect linear model with nested classification. *Scandinavian Actuarial Journal*, 1986(3-4):204–222.

Oh, R., Ahn, J. Y., and Lee, W. (2019). On copula-based collective risk models. *arXiv, https://arxiv.org/abs/1906.03604*.

Oh, R., Shi, P., and Ahn, J. Y. (2020). Bonus-malus premiums under the dependent frequency-severity modeling. *Scandinavian Actuarial Journal*, 2020(3):172–195.

Park, S. C., Kim, J. H., and Ahn, J. Y. (2018). Does hunger for bonuses drive the dependence between claim frequency and severity? *Insurance: Mathematics and Economics*, 83:32–46.

Peng, F. and Dey, D. K. (1995). Bayesian analysis of outlier problems using divergence measures. *Canadian Journal of Statistics*, 23(2):199–213.

Pinquet, J. (1997). Allowance for cost of claims in bonus-malus systems. *ASTIN Bulletin: The Journal of the IAA*, 27(1):33–57.

Pinquet, J. (1998). Designing optimal bonus-malus systems from different types of claims. *ASTIN Bulletin: The Journal of the IAA*, 28(2):205–220.

Pinquet, J., Guillén, M., and Bolancé, C. (2001). Allowance for the age of claims in bonus-malus systems. *ASTIN Bulletin: The Journal of the IAA*, 31(2):337–348.

Rootzén, H., Tajvidi, N., et al. (2006). Multivariate generalized pareto distributions. *Bernoulli*, 12(5):917–930.

Salvadori, G. and De Michele, C. (2007). On the use of copulas in hydrology: theory and practice. *Journal of Hydrologic Engineering*, 12(4):369–380.

Sánchez-Sánchez, M., Sordo, M., Suárez-Llorens, A., and Gómez-Déniz, E. (2019). Deriving robust bayesian premiums under bands of prior distributions with applications. *ASTIN Bulletin: The Journal of the IAA*, 49(1):147–168.

Shevchenko, P. and Wuthrich, M. (2006). The structural modelling of operational risk via bayesian inference: Combining loss data with expert opinions.

Shi, P., Feng, X., and Ivantsova, A. (2015). Dependent frequency-severity modeling of insurance claims. *Insurance: Mathematics and Economics*, 64:417 – 428.

Shi, P. and Valdez, E. A. (2012). Longitudinal modeling of insurance claim counts using jitters. *Scandinavian Actuarial Journal*, 2012:1 – 21.

Shi, P. and Valdez, E. A. (2014). Multivariate negative binomial models for insurance claim counts. *Insurance: Mathematics and Economics*, 55:18–29.

Shi, P. and Yang, L. (2018). Pair copula constructions for insurance experience rating. *Journal of the American Statistical Association*, 113(521):122–133.

Shih, J. H. and Louis, T. A. (1995). Inferences on the association parameter in copula models for bivariate survival data. *Biometrics*, 51(4):1384–1399.

Sklar, A. (1959). Fonctions de repartition a n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Universite de Paris*, (8):229–231.

Torre, R., Myrskylä, M., et al. (2011). *Income inequality and population health: a panel data analysis on 21 developed countries*. Max-Planck-Institut für demografische Forschung.

Winkelmann, R. (2008). *Econometric analysis of count data*. Springer Science & Business Media.

Yang, L., Frees, E. W., and Zhang, Z. (2019). Nonparametric estimation of copula regression models with discrete outcomes. *Journal of the American Statistical Association*, pages 1–25.

Yang, X., Frees, E. W., and Zhang, Z. (2011). A generalized beta copula with applications in modeling multivariate long-tailed data. *Insurance: Mathematics and Economics*, 49(2):265–284.

Young, V. R. (1998). Robust bayesian credibility using semiparametric models. *ASTIN Bulletin: The Journal of the IAA*, 28(2):187–203.

Zhang, T. (2004). Statistical behavior and consistency of classification methods based on convex risk minimization. *Annals of Statistics*, pages 56–85.